Data and outcomes archiving: issues and opportunities

Joe Deasy
Division of Bioinformatics and Outcomes Research,
Dept. of Radiation Oncology

Outline

- Current 3D de-archiving capabilities: the view of an outside de-archiver
- Current 3D & outcomes RTOG de-archiving capabilities: the view of an outcomes modeler
- Future opportunities: the databank paradigm

The view from a outside de-archiver: user needs

- Timely de-archiving (i.e., need capacity to accommodate requests)
- Convenient data media
- Thoroughly QA'd plans (i.e., data is reliable)
- Consistent dose algorithms
- Accurate dose algorithms
- Image and structure sets are needed

3D de-archiving capabilities: the view of an outside de-archiver

- Variations in dose calculation algorithms
  - Is water-based good enough?
  - Heterogeneity corrections are desirable if ‘good’ (old heterogeneity corrections are not clearly better than water-based, at least for lung)
  - Consistent dose calc quality is very desirable

De-archiving: the view from an outcomes modeler (1/2)

- Outcomes data is easier to handle
- Need very large datasets to drive sound or novel statistical observations
- The time-parameters for treatments should always be clear
- Txs which are a compound of more than one dose-distribution should store multiple dose distributions separately, to allow for modeling of fractionation effects

De-archiving: the view from an outcomes modeler (2/2)

- Dose-volume outcomes modeling is quite different (though not completely different) from traditional RTOG data analyses
- Some disconnect for this type of analysis with the RTOG statistical staff
- Multiple types of analyses are possible

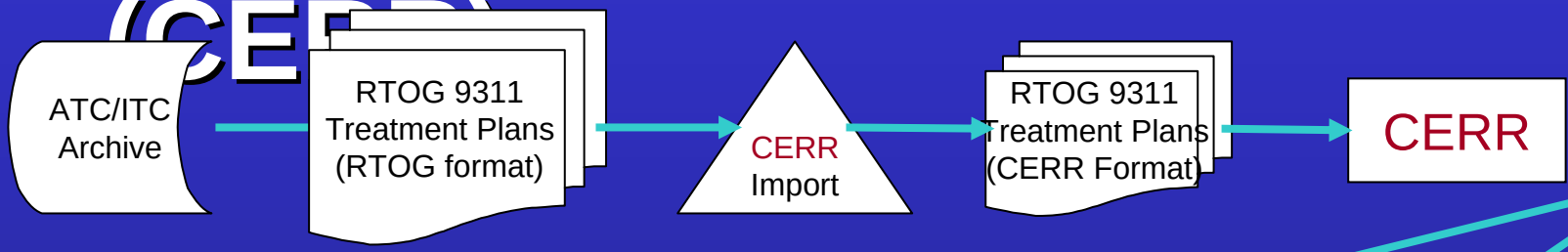
- Multiple types of analyses are possible, so the same dataset should be made available to multiple groups

What data is needed? Almost all of it…

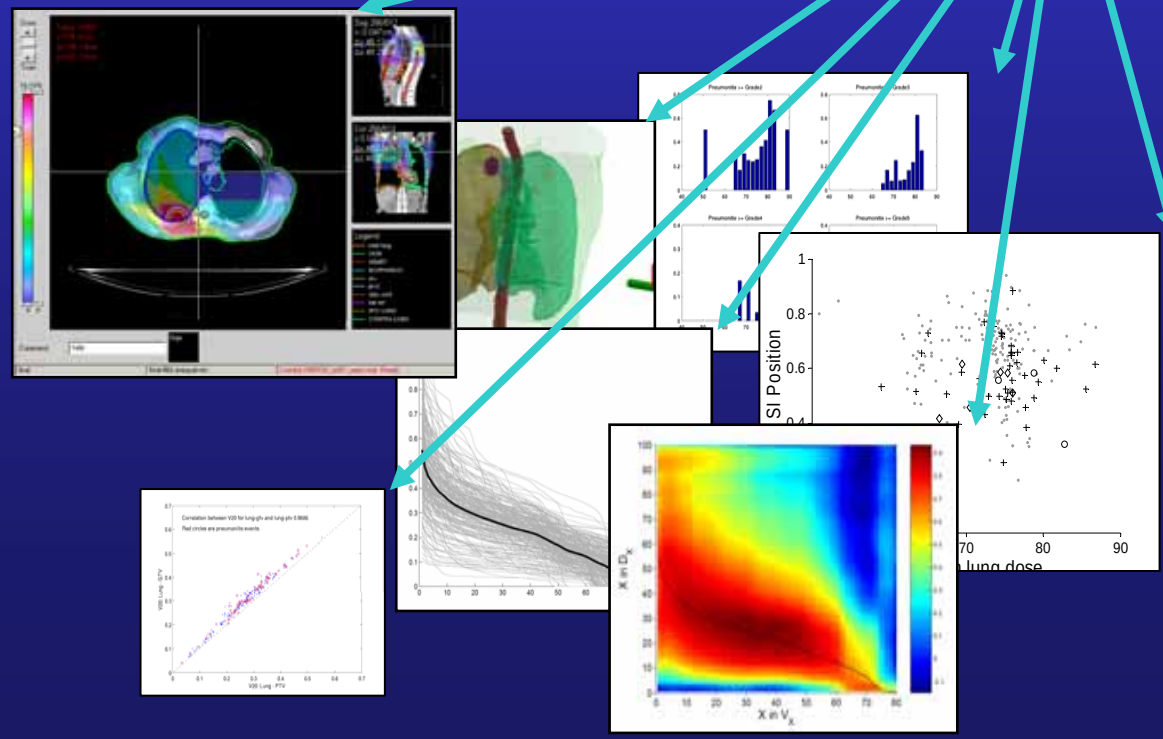
- New metrics of importance might be based on:
  - Anatomical position (e.g., distance from PTV to the spinal cord)
  - Image-weighted (i.e., CT-weighted lung DVHs)
  - Positional variation-weighted
    - E.g., convolution/Monte Carlo of rectal dose over position
  - Rectal filled vs. not filled (e.g., MDACC prostate outcomes correlation)
  - Image values (CT, PET, MRI),
  - contoured structures, dose distributions
  - Type of CT scan (slow vs. fast)
  - 'Corrected' images/dose values, e.g., corrected for breathing motion
  - DVHs
  - ?

Environment for Radiotherapy Research (CERR)

Radiation Therapy Oncology Group



- **Custom software extracts dose, volume, and structure data automatically**
- **http://radium.wustl.edu/cerr**



De-archiving: why is a graphically-based data-mining tool necessary?

- User's must be able to review plans in a treatment planning-like graphical fashion
  - To verify correctness
  - To isolate potentially important plan characteristics
- User's must be able to conveniently extract datasets
  - No one has the time to do this except in batch mode
  - The metrics which can be extracted will vary widely

The view from an outside de-archiver

- Conversion into user-friendly format would be desirable (i.e., CERR/Matlab). Why?
  - ITC could QA the conversion process
  - ITC could compare with previously derived metrics
  - I.e., ITC could offer good assurance that the data was correct and usable when it left ITC
  - Oddball processing issues (e.g., feet-in-first scan data, unexpected DICOM tags, etc.)
  - Obviously a lot of extra work for ITC…

Predictors of Lung Toxicity from the RTOG 9311 Radiation Dose Escalation Trial; Dose / Volume and GTV Position are Important

Radiation Therapy Oncology Group

RTOG Secondary Analysis

Jeffrey Bradley, Joseph Deasy, Andrew Hope, Patricia Lindsay, Issam El Naqa, Walter Bosch, John Matthews, William Sause, and Mary Graham

Objectives

Radiation Therapy Oncology Group

- To determine the predictors of radiation pneumonitis for patients enrolled on RTOG 9311
- To test our institutional model for predicting radiation pneumonitis with the 9311 dataset

RTOG 9311 Dataset

Radiation Therapy Oncology Group

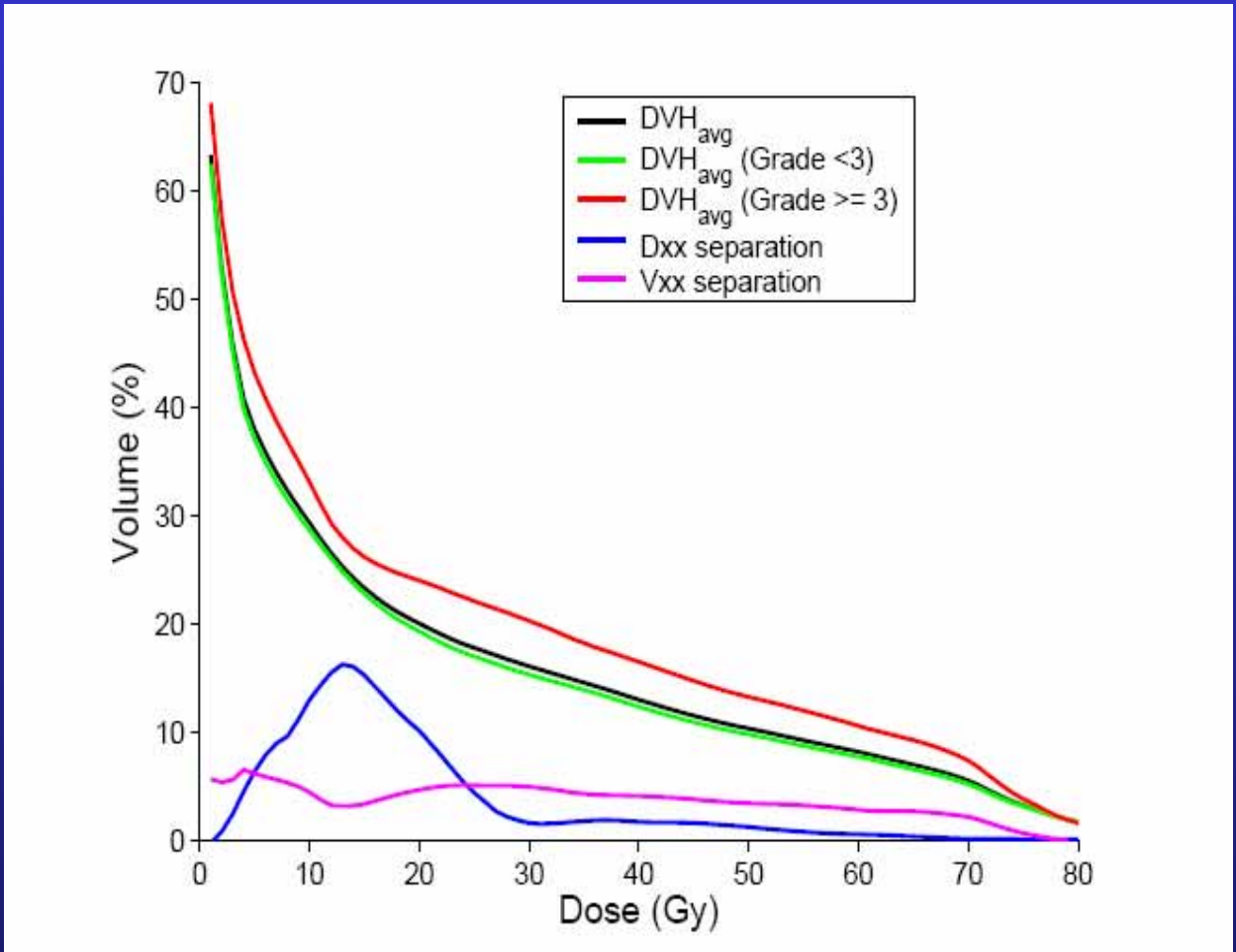
- 179 patients enrolled on RTOG 9311
  - 10 plans were incomplete submissions to ITC
  - 31 had <6 months clinical follow up
  - 9 had missing data points
- 129 patients evaluated

Parameters Assessed

Dosimetric / Geometric	Patient
• Mean lung dose • GEUD • GTV volume • GTV center position • Dx and Vx parameters at 5 Gy steps • Maximum dose	• Age • Sex • Performance status • Pre-RT chemo • Pre-RT FEV1 • Pre-RT DLCO • Pre-RT hemoglobin

RTOG 9311 Lung DVH

Radiation Therapy Oncology Group



- **D15 shows most separation**

Multi-metric Modeling

Radiation Therapy Oncology Group

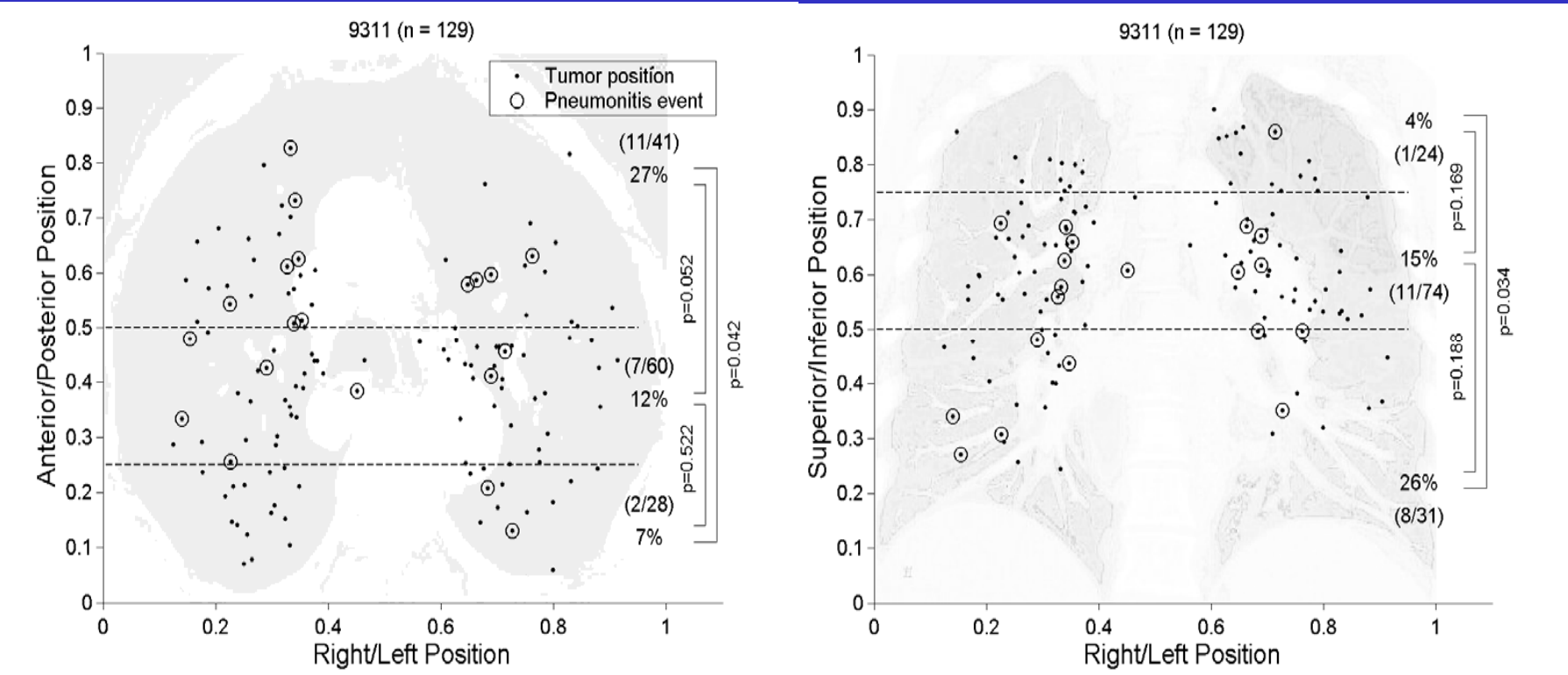
Logistic regression to fixed model order on multiple bootstrap samples

Model order	Most frequent model parameters
1*	D15
2	D15, GTV-AP
3	D15, GTV-AP, GTV-SI
4	D15, D25, GTV-AP, GTV-SI
5*	D15, D25, D45, GTV-AP, GTV-SI

*** Models with highest predictive power from simulations**

GTV Position

Radiation Therapy Oncology Group



- **RTOG 9311 shows differences in AP and SI position**

Conclusions

Radiation Therapy Oncology Group

- Dose / volume and GTV position parameters are important for predicting RP (specifically Dx, AP and SI position)
- Dx values (i.e. D15) are dataset specific and may not be generalizeable across institutions
- Future plans are to build predictive models using multi-institutional datasets

Overall experience was excellent

What could make it better?

- Public availability of data for others, to be combined/tested with future datasets and models
- Accurate dose calculations
  - Difficult for RTOG format archives
  - Practical for DICOM files

The big opportunity: databank

- Would allow single-institution or collaborative trial investigators to
  - Archive high quality outcomes and treatment planning data
  - Test and compare treatment results against growing database
  - Improve models of tumor and normal tissue response to radiation
  - Learn much more effectively from the past
  - Make RT meta-analyses much more precise

The big opportunity: databank

- Should be facilitated by a respected and competent umbrella organization
- Could utilize growing QA and informatics capabilities of ATC
- Would be an attractive initiative to NCI/NIH in view of stated goals of creating user-accessible databases of NCI funded results
- Would include outcomes

Other opportunities

- Facilitate greater user access and manipulation of data via CERR/Matlab
- Facilitate greater dosimetric accuracy via Monte Carlo recalculations of DICOM submissions