Imaging Guided Intervention Trials: Concepts and Challenges

Michael C. Soulen, MD FSIR FCIRSE
Professor of Radiology and Surgery
Abramson Cancer Center
University of Pennsylvania
Plumbing Trials: Goals & Outcomes

- **PAD/PVD**
  - Limb salvage
  - Wound healing
  - Symptom relief
  - Venous obstruction or insufficiency

- **TIPS**
  - Ascites/Bleeding

- **Dialysis & Venous Access**

- **Non-vascular**
  - GI/GU/Biliary

- **Clinical Endpoint**

- **Primary patency**
  - Plumbing open without intervention

- **Primary Assisted Patency**
  - Plumbing kept open with additional intervention

- **Secondary Patency**
  - Plumbing occluded and restored
IO Trials: Goals & Outcomes

GOAL
- Cure
- Prevention of progression as the cause of death
- Surrogate Endpoint: disease control (CR+PR+SD)
- Surrogate Endpoint: response (CR+PR)
- Palliation

OUTCOME
- Survival (OS, DFS, CSS)
- Survival
- PFS; TTP
- Response
- Pain, function, QOL
Survival

- It’s what patients care about
- Essential component of informed consent
- The gold standard for Phase III clinical trials
- FDA requires it
- Longest time (5+ years)
- Largest sample size (hundreds)
- Most expensive to measure (tens of millions)
What is Survival? - not so simple!

- Overall Survival
  - Alive or Dead
  - Appropriate for aggressive diseases where death from cancer is the expected outcome (lung, pancreatic)

Lung Cancer 5-year Survival

What is Survival? - not so simple!

- **Disease-Free Survival**
  - Alive without cancer
  - Appropriate for less aggressive diseases where prolonged remission and survival after recurrence is expected (breast, RCC, NET)
  - Death from cancer still predominates, but OS not a practical primary endpoint

- **Cancer-Specific Survival**
  - Death from cancer
  - Appropriate for diseases where non-cancer related death predominates (prostate, T1 renal cell)
  - Effect measured by Cumulative Incidence Estimate rather than Kaplan-Meier, compare with Gray’s Test instead of log rank test
Example: Cancer-Specific Survival

- 70 year old with T1 RCC
- What is the probability of death in 5 years?
- What is the risk of death from RCC in 5 years?
- 5 year cancer-specific survival for untreated tumors < 4cm (T1a) is 100%, >90% for T1b.
- Rate of metastasis <1%

Prostate Cancer: cancer-specific and overall survival by age stratum.

Bechis S K et al. JCO 2011;29:235-241
How do we measure survival benefit?

**Overall survival**

- **Sorafenib (n=299) = 10.7 months**
- **Placebo (n=303) = 7.9 months**

**Time-to-progression**

- **Sorafenib (n=299) = 5.5 months**
- **Placebo (n=303) = 2.8 months**

**HR = 0.69**

(95% CI: 0.55 – 0.87; p < 0.001)

**HR = 0.58**

(95% CI: 0.45 – 0.74; p < 0.001)

HR = hazard ratio
CI = confidence interval

Survival: Hazard Ratio vs. Median

- No one has the median survival - could be months less but **years** more
- Medians mask subpopulation effects
- Should use Hazard Ratios

Ruutiainen AT JVIR 2007;18:847-55
Kaplan Meier survival curve for chemoembolization of sarcoma metastases
So what have we learned?

- “Survival” is usually the most important and rigorous primary outcome
- Be sure you know which type of “survival” matters
- Measure with Kaplan-Meier estimates
- Compare with Hazard Ratios

- but it’s not that simple....
IO Trial Design Conundrums

Time-based Outcomes (OS, PFS, TTP):

- for systemic therapy trials, clock starts with initiation of the drug, ends with progression of disease anywhere or death

- IO therapies may be staged: not all tumor treated at same time

- IO therapies may be repeatable: first progression may not signal failure of therapy
Staged Therapies

<table>
<thead>
<tr>
<th>Month</th>
<th>Baseline Imaging</th>
<th>embo 1</th>
<th>embo 2</th>
<th>embo 3</th>
<th>embo 4</th>
<th>Assessment Imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Trial Design for Staged Therapies

ECOG 1208: TACE +/- sorafenib

**Schema**

- **Arm A**
  - Sorafenib + Chemoembolization
  - Sorafenib: 400 mg (2 x 200 mg tablets) po b.i.d. until progression
  - Chemoembolization: up to 4 Chemoembolizations
  - To be completed within 6 months

- **Arm B**
  - Placebo + Chemoembolization
  - Placebo: 2 tablets po b.i.d. until progression
  - Chemoembolization: up to 4 Chemoembolizations
  - To be completed within 6 months

**Stratification**
- Factor 1: Macrovascular intrahepatic portal vein invasion present vs. Macrovascular intrahepatic portal vein invasion absent
- Factor 2: Child Pugh A vs. Child Pugh B7

**Endpoints**
- Primary: FFS
- Secondary: Safety Correlative Studies

**Randomize**
Trial Design for Staged Therapies

ECOG 1208: TACE +/- sorafenib
Scheduled Imaging

5.3.7 Tumor imaging studies for the purposes of determining PFS are to be performed as follows:

- Baseline (chest CT, Abdomen/pelvis CT or MRI)
- 4 months after first chemoembolization (chest CT, Abdomen/pelvis CT or MRI)
- 8 months after first chemoembolization (chest CT, Abdomen/pelvis CT or MRI)
- Every 2 months beginning at 10 months post baseline
**INCLUSION CRITERIA**
- Unresectable, multinodular HCC
- Child-Pugh A without ascites or encephalopathy
- ECOG PS of 0

**EXCLUSION CRITERIA**
- Vascular invasion (VI)
- Extrahep. spread (EHS)
- Prior TACE, prior systemic therapy

**TACE (optional)**
- Cycle no (=4 weeks)
- First DEB-TACE performed 3-7 days after start of sorafenib or placebo
- Subsequent DEB-TACE performed on day 1 (±4 days) of cycles 3, 7, and 13, and every 6 cycles thereafter
- Patients allowed optional DEB-TACE sessions between cycles 7-13 and 13-19

**Primary**
- TTP (central review)

**Secondary**
- Time to VI/EHS
- Overall survival
- Safety
- Others

*Lencioni R et al. ASCO GI 2012*
EPOCH: required single-session whole-liver therapy

- Colorectal metastases progressing on 1st-line systemic chemotherapy
- Randomized to 2nd-line chemotherapy +/- Y90 radioembolization
- Requires whole-liver radioembolization
Disease Control

- **Time-to-Progression (TTP) or Progression-Free Survival (PFS)**
  - proposed endpoint for Phase II trials
  - faster and cheaper than survival

- **Limitations:**
  - doesn’t necessarily correlate with survival
  - problematic to measure with staged therapies
  - relevance if therapy is repeatable (ablation, embo)?
  - vulnerable to high censorship – OLT, non-cancer death (RCC, HCC, prostate)
  - can be surprisingly hard to measure (RADIANT-2)
FDA Advisory Committee
evirolimus vs. placebo for carcinoid

Discordant Results

Carcinoid Trial (C2325) 2nd interim analysis demonstrated discordance of PFS between INV and IRC

- IRC PFS ($p=0.233$) crossed the Futility Boundary ($p = 0.175$)
- INV PFS ($p=0.003$) crossed the Efficacy Boundary ($p = 0.010$)
TTP Limitations

TTP Hazard Ratio vs. Median

Same issues as for survival

Ruutilainen AT JVIR 2007;18:847-55
## Repeatable Therapies

<table>
<thead>
<tr>
<th>TIME</th>
<th>PATIENT 1</th>
<th>PATIENT 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>A0 ablation</td>
<td>A0 ablation</td>
</tr>
<tr>
<td>6 months</td>
<td>(-) recurrence</td>
<td>(+) recurrence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A0 ablation</td>
</tr>
<tr>
<td>12 months</td>
<td>(-) recurrence</td>
<td>(-) recurrence</td>
</tr>
<tr>
<td>PFS</td>
<td>12 months</td>
<td>6 months</td>
</tr>
<tr>
<td>DFS/CSS/OS</td>
<td>12 months</td>
<td>12 months</td>
</tr>
</tbody>
</table>
# Trial Design for Repeatable Therapy

## Specific Aims/Objectives

The primary aim for this trial is to estimate the proportion of patients undergoing solitary or repetitive percutaneous RFA treatment sessions whose livers have no identifiable tumor by CT scan at 18 months following initiation of therapy.

Since there is no direct comparison rate in this study, it was sized to detect a difference between a disappointing and promising rate. Based on the literature, the surgical success rate is approximately 65% @ 18 months. It is believed that RFA could have a considerably better success rate of approximately 85%. A sample size of 35 participants will provide at least 85% power to detect the proposed alternative RFA success rate (assuming a null hypothesis of 65%)

### Multicenter Feasibility Study of Percutaneous Radiofrequency Ablation of Hepatocellular Carcinoma in Cirrhotic Patients

**Schema**

<table>
<thead>
<tr>
<th>E</th>
<th>Cirrhosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>1-3 HCC ≤ 3.0 cm or</td>
</tr>
<tr>
<td>G</td>
<td>A single tumor &gt;3.0 cm but ≤ 5cm</td>
</tr>
<tr>
<td>I</td>
<td>(Repeat RFA permitted for)</td>
</tr>
<tr>
<td>T</td>
<td>Initial ablation session)</td>
</tr>
<tr>
<td>R</td>
<td>Percutaneous RFA</td>
</tr>
<tr>
<td>M</td>
<td>CT scans within 1st week</td>
</tr>
<tr>
<td>L</td>
<td>after initial RFA and</td>
</tr>
<tr>
<td>L</td>
<td>every 3 months post ablation</td>
</tr>
<tr>
<td>O</td>
<td>for 18 months*</td>
</tr>
<tr>
<td>W</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
</tr>
</tbody>
</table>
Trial Design for Repeatable Therapy

- **Time-to-Treatment-Failure (TTTF)**
- **Time to Untreatable Progression (TTUP)**
  - Progression when test therapy can no longer be performed
  - pre-determined objective criteria for untreatable progression
  - technical (can’t get there) vs. clinical (declining liver function or performance status) criteria
  - who decides?
  - are criteria reproducible?
  - appealing concept but hard to do
Response: the 4 Essentials

- **Accuracy** -- what you measure is truth
- **Precision** -- you and everyone else get the same measurement, every time
- **Simple & Generalizable** -- anyone can do it in daily practice without limiting workflow
- **Useful** -- informs management decision or prognosis
Response Examples

- RECIST/WHO
- Necrosis ("EASL")
- mRECIST
- Lipiodol retention
- Functional imaging
  - PET, diffusion, etc.
RECIST 1.1
What Med Oncs Use

- Sum of single longest diameters of index tumors (2 per organ, up to 5)
- CR = gone
- PR = 30% reduction in sum of LD’s
- Progression = 20% increase in sum of LD’s or new tumors
- Stable = neither PR or progression
What is the scientific basis for RECIST?

- -30% and +20% diameter change is equivalent to a 50% change in tumor volume
- -30% and +20% diameter changes statistically correlate with patient survival
- Single linear diameters are more precise and accurate than bidimensional (WHO) measurements
- A bunch of medical oncologists palpating a foam mattress
History of response criteria

Rubber Foam

Soft Mattress

Solid Spheres (1.8-14.5 cm)

Moertel and Hanley, Cancer (1976) 38:388-394
History of response criteria

• Sixteen oncologists determined the diameter of 12 spheres (1.8-14.5 cm)

• The measured size of identical spheres differed
  – by at least 25% in 25% of the measurements
  – by at least 50% in 6.8% of the measurements

("false-positive rate for response")

Moertel and Hanley, Cancer (1976) 38:388-394
2 cm lung nodule

3 radiologists re-reading same scan and scan done 15 min later

% difference in measurement

<table>
<thead>
<tr>
<th>Measurement, Comparison, and Reader</th>
<th>Tumor Size on Scan 1</th>
<th>±2 cSDs</th>
<th>Difference*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unidimensional (cm)</td>
<td>2.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scan 1 vs scan 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.66, 2.40</td>
<td>-16.8, 20.1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.59, 2.52</td>
<td>-20.6, 26.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.59, 2.51</td>
<td>-20.4, 25.6</td>
<td></td>
</tr>
<tr>
<td>Scan 1 repeat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 repeat</td>
<td>1.81, 2.21</td>
<td>-9.5, 10.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.65, 2.42</td>
<td>-17.4, 21.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.57, 2.55</td>
<td>-21.6, 27.5</td>
<td></td>
</tr>
</tbody>
</table>

20%-30% difference in measuring same lesion on same scan or concurrent scan by same reader

Zhao, Radiology 2009;252:263-272
RECIST: Why -30%/+20%?

Practical limitation of ACCURACY and PRECISION

NOT based on clinical outcomes

<table>
<thead>
<tr>
<th></th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>scan 1</td>
<td>7</td>
<td>21</td>
<td>45</td>
<td>73</td>
</tr>
<tr>
<td>scan 2</td>
<td>10</td>
<td>25</td>
<td>50</td>
<td>85</td>
</tr>
</tbody>
</table>

- +16% = Stable Disease
- DO NOT report this as progression
  - med oncs don’t look at pictures
  - implies treatment failure - alters therapy
  - you make patients cry
Response vs. Survival

Colon Cancer: chemoembolization

Partial response, died 6 months
Colon Cancer: chemoembolization

Stable disease at 1 year, CEA down 90%, ECOG 0
Chemoembolization of HCC: RECIST

Survival by RECIST Response

Survival probability (%)

months

<table>
<thead>
<tr>
<th>RECIST_CAT</th>
<th>N</th>
<th>%</th>
<th>2 YR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>8</td>
<td>11</td>
<td>61%</td>
</tr>
<tr>
<td>SD</td>
<td>50</td>
<td>69</td>
<td>65%</td>
</tr>
<tr>
<td>PD</td>
<td>10</td>
<td>14</td>
<td>64%</td>
</tr>
</tbody>
</table>

P=0.55
Post RF Ablation: It Grew!

- Contrast enhanced CT
  - Pre RFA
- Contrast enhanced CT
  - 6 month post RFA
The “EASL Criteria” Myth

- EASL 2000 Consensus Document
- “should take tumor necrosis into account”
- no criteria!!!

Bruix, J Hepatol 2001;35:421-30
Necrosis Response Criteria

mRECIST

- Single longest diameter of enhancing (viable) tumor
- CR = none, PR 30% reduction from baseline, PD 20% increase from baseline or new lesions

Lencioni, Semin Liver Dis 2010;30:52-60
Chemoembolization of HCC: mRECIST

Survival by mRECIST

<table>
<thead>
<tr>
<th>CR</th>
<th>44</th>
<th>61%</th>
<th>69%</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR</td>
<td>14</td>
<td>19%</td>
<td>51%</td>
</tr>
<tr>
<td>SD</td>
<td>11</td>
<td>15%</td>
<td>41%</td>
</tr>
<tr>
<td>PD</td>
<td>3</td>
<td>4%</td>
<td>-</td>
</tr>
</tbody>
</table>

P=0.47
Chemoembolization of HCC: mRECIST reader confidence

60% of measurements of residual viable tumor diameter classified as “not confident”
## NWU Necrosis Criteria

<table>
<thead>
<tr>
<th>Response Type</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Response (CR)</td>
<td>100% decrease in amount of enhancing tissue in index lesion</td>
</tr>
<tr>
<td>Partial Response (PR)</td>
<td>≥50% decrease in amount of enhancing tissue in index lesion</td>
</tr>
<tr>
<td>Stable Disease (SD)</td>
<td>&lt;50% decrease in to ≤25% increase in amount of enhancing tissue in index lesion</td>
</tr>
<tr>
<td>Progressive Disease (PD)</td>
<td>&gt;25% increase in amount of enhancing tissue in index lesion&lt;br&gt;New lesions or metastases&lt;br&gt;New vascular invasion&lt;br&gt;New enhancement in previously treated index lesion warranting further locoregional therapy</td>
</tr>
</tbody>
</table>

**Simple visual estimate**
## NWU Necrosis Criteria

<table>
<thead>
<tr>
<th></th>
<th>6 mo</th>
<th>12 mo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>SD</td>
</tr>
<tr>
<td>N</td>
<td>42</td>
<td>28</td>
</tr>
<tr>
<td>Survival</td>
<td>25</td>
<td>14</td>
</tr>
</tbody>
</table>

p = 0.002  
p = 0.0001

Memon K, Gastro 2011 epub
U Penn Necrosis Criteria

Survival by % Necrosis

P=0.07

<table>
<thead>
<tr>
<th>Quintiles</th>
<th>N</th>
<th>%</th>
<th>2 YR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>75</td>
<td>69%</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>11</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>6</td>
<td>50%</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>33%</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>4</td>
<td>0%</td>
</tr>
</tbody>
</table>
Shim RAD 2012;262;708-718

Pre-TACE

Post-TACE

Percent change

1. \( \frac{(A' + a')}{(A + a)} \times 100 \text{ (\%)} \) : by RECIST

1. \( \frac{(A'' + a'')}{(A + a)} \times 100 \text{ (\%)} \) : by mRECIST

1. \( \frac{(a' \times b') + (a'' \times b'')}{(a \times b) - (a \times b)} \times 100 \text{ (\%)} \) : by WHO

1. \( \frac{(a' \times b') + (a'' \times b'')}{(a \times b) - (a \times b)} \times 100 \text{ (\%)} \) : by EASL
Precision: single ROI, mult ROI, volumetric ROI
Precision: single ROI, mult ROI, volumetric ROI

Precision was 12.0% for the single-ROI method
9.3% for the average-measurement-in-three-ROIs method
3.3% for the volumetric method.

Chalian, RAD 2012;262:853
Radiologic-Pathologic Analysis of contrast-enhanced and Diffusion-weighted MR imaging in Patients with HCC after TACE: Diagnostic Accuracy of 3D Quantitative Image Analysis

17 HCC resected after TACE

3D segmentation for contrast enhancement and ADC

Slice-by-slice correlation with histology
Old Concept: Lipiodol as an imaging biomarker for response

Path validation

Oil Retention vs. Survival

Radiological response to TACE predicts HCC survival.

(a) Compact lipiodolisation vs. Noncompact lipiodolisation

(b) Child-Pugh class A vs. Child-Pugh class B

(c) TNM stage II vs. TNM stage III

Kim, AP&T 2012;35:1343
New Concept: Index lesion

WHO Criteria

NWU Necrosis Criteria
New Concept: PET post Y-90

- Stable on CT
- Improvement on PET

courtesy R. Salem
PET post Y-90

- Stable on CT

- Resolution on PET
**Y-90 PET vs. CT**

<table>
<thead>
<tr>
<th>Post-treatment PET</th>
<th>Response</th>
<th>No Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>No Response</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Take Home Points

• Image-guided therapies pose multiple challenges in clinical trial design
• Time-based outcomes such as survival and disease control should allow for staged and repeated therapy
• Imaging-based endpoints surprisingly difficult to standardize
Take Home Points

- RECIST not useful for embolization and ablation
- mRECIST better
- A simple visual volumetric estimate of necrosis is promising -- needs validation of reliability and scale
- Potential for automated CAD of necrosis volume with greater reliability than human reads
- Other functional readouts under investigation - diffusion, metabolism
- Concept of the index treated lesion may compensate for staged therapies, if validated