The need for QI in Medical Oncology Treatment Guidance and Assessment

Robert Jeraj
Professor of Medical Physics, Human Oncology, Radiology and Biomedical Engineering
Director of Translational Imaging Research Program
University of Wisconsin Carbone Cancer Center, Madison, WI
University of Ljubljana, Slovenia
rjeraj@wisc.edu

Disclosures

- AIQ Solutions, Founder and President

Types of imaging

Qualitative imaging (Diagnostics)
Quantitative imaging (Quantitative Imaging Biomarkers)
**Biomarkers and surrogate endpoints**

- **Biomarkers** are characteristics that are objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

- **Biomarkers as surrogate endpoints** are biomarkers that are intended to substitute for clinical endpoints. Surrogate endpoints are expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.

**Types of biomarkers**

- Risk assessment
- Early detection
- Prognostic
- Predictive
- Treatment response
- Recurrence

**Timeline**

- **Before**
  - Diagnosis and staging
  - Target definition

- **During**
  - Early treatment assessment

- **After therapy**
  - Late treatment assessment
What does make imaging biomarkers QUANTITATIVE?


Repeatability and reproducibility

- **Repeatability**: The ability of the QIB to repeatedly measure the same feature under identical or near-identical conditions.
  - These studies are often referred to as test–retest, scan–rescan, or coffee-break experiments

- **Reproducibility**: The reliability of the QIB measuring system in different conditions that might be expected in a preclinical study or clinical trial or in clinical practice (e.g. multiple sites)


TREATMENT RESPONSE ASSESSMENT
Response assessment today

- **WHO** (1979, 1981)\(^1,2\)
  - anatomic
  - bidimensional measurement of lesion

- **RECIST** (2000, 2009)\(^3,4\) – *Response Evaluation Criteria in Solid Tumors*
  - anatomic
  - CT/MR based
  - unidimensional measurement of lesion
  - 4 response categories (CR, PR, SD, PD)
    - Complete Response: disappearance
    - Partial Response: >30% decrease
    - Stable Disease: in between
    - Progressive Disease: >20% increase

Anatomical vs functional response

- **PET-based response assessment**
  - **EORTC, NCI Recommendations** (1999, 2005)\(^1,2\)
    - SUV-based approach
    - SUV\(_{\text{mean}}\) and SUV\(_{\text{max}}\)
    - Response categories with thresholds (CR, PR, SD, PD)
    - Problems
      - SUV\(_{\text{mean}}\): collapse information, sensitivity issues
      - SUV\(_{\text{max}}\): noise contamination
      - fails to use all available functional data
      - heterogeneity
      - no response threshold validation
      - few sensitivity studies
      - alternative measures

- **PET Response Criteria in Solid Tumors (PERCIST)** (2009)\(^3\)
  - SUV\(_{\text{peak}}\)

\(^1\) WHO handbook 1979
\(^2\) Miller et al. 1981
\(^3\) Therasse et al. 2000
\(^4\) Eisenhauer et al. 2009

\(^1\) Young et al. 1999
\(^2\) Shankur et al. 2006
\(^3\) Wahl et al. 2009
What about harmonization?

Lung Lesion

<table>
<thead>
<tr>
<th>Month</th>
<th>Original</th>
<th>Harmonized</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>D710</td>
<td>D710</td>
</tr>
<tr>
<td>6</td>
<td>DVCT</td>
<td>D710</td>
</tr>
<tr>
<td>12</td>
<td>DVCT</td>
<td>D710</td>
</tr>
</tbody>
</table>

Maximum SUV

METASTATIC PROSTATE CANCER
Metastatic disease

- **Metastasis** (μετά ("next") + στάσις ("placement")

Cancer-related deaths

- **Bone metastases** present in 90% of advanced prostate cancer patients (Bubendorf et al, Hum Pathol. 2000)
  - Relative 5-year survival rate is about 30% (SEER 2006-2012)

Common destinations from solid tumors: bones, brain, lungs, liver

How can molecular imaging help?

- **Whole body imaging** provides a complete snapshot of the progression of multiple metastases
- The role of imaging in monitoring treatment response is not well established (Wallace et al, J Cancer 2014)
  - Often based on responses of a few lesions
  - Limited assessment in disease burden

- **18F-NaF PET/CT**, imaging of bone metabolism, ideal for monitoring bone metastases
  - 100% sensitivity and 70-100% specificity (Sapir et al, J Nucl Med 2006)
  - Quantitative and reproducible whole body scans (Kurdziel et al, J Nucl Med 2012)

Performance characteristics of QIB
Quantitative Total Bone Imaging (QTBI)

Scan 1
NaF PET/CT Scan
Lesion segmentation
Image feature quantification

Scan 2
NaF PET/CT Scan
Lesion segmentation
Image feature quantification

Registration
Lesion matching
Difference quantification


Quantitative Total Bone Imaging (QTBI)

NaF PET/CT acquisition
Uptake Localization
Lesion Segmentation

SUV

Quantitative Total Bone Imaging (QTBI)

NaF PET/CT #1
NaF PET/CT #2
Articulated Registration
Lesion-based Response

SUV

Progressing
Stable
Responding
Multicenter trial of metastatic castrate-resistant prostate cancer patients – received pre-treatment test-retest 18F-NaF PET/CT scans

<table>
<thead>
<tr>
<th>Site</th>
<th>Patients</th>
<th>Bone lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Wisconsin Carbone Cancer Center</td>
<td>18</td>
<td>265</td>
</tr>
<tr>
<td>Memorial Sloan-Kettering Cancer Center</td>
<td>11</td>
<td>78</td>
</tr>
<tr>
<td>National Cancer Institute</td>
<td>6</td>
<td>68</td>
</tr>
<tr>
<td>All</td>
<td>35</td>
<td>411</td>
</tr>
</tbody>
</table>

Test/retest scans (3-5 days apart)

Standardized Uptake Value (SUV) metrics extracted from an ROI

SUV\text{max} – maximum uptake
SUV\text{mean} – average uptake
SUV\text{total} – total uptake

Repeatability of QTBI (NaF PET/CT)

<table>
<thead>
<tr>
<th>Feature</th>
<th>SUV\text{max}</th>
<th>SUV\text{mean}</th>
<th>SUV\text{total}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>48.2</td>
<td>28.8</td>
<td>19.4</td>
</tr>
<tr>
<td>Retest</td>
<td>48.3</td>
<td>19.4</td>
<td>92.7</td>
</tr>
</tbody>
</table>

SUV\text{max} – maximum uptake
SUV\text{mean} – average uptake
SUV\text{total} – total uptake

What is our quantitative accuracy?

Low repeatability

Test:
- SUV
- Feature
- SUV\text{max}
- SUV\text{mean}
- SUV\text{total}

Retest:
- SUV
- Feature
- SUV\text{max}
- SUV\text{mean}
- SUV\text{total}

High repeatability

Test:
- SUV
- Feature
- SUV\text{max}
- SUV\text{mean}
- SUV\text{total}

Retest:
- SUV
- Feature
- SUV\text{max}
- SUV\text{mean}
- SUV\text{total}
**How much is repeatability?**

Bland-Altman plots assess measurement agreement (Bland and Altman, J Bioph Stat 2007).

UWCCC significantly narrower variance (p < 0.001)

**Defining limits of agreement**

Test-retest limits of agreement (LOA) can be used to define significant changes in imaging features in individual lesions (Lin et al., J Nucl Med 2016).

**Imaging site affects repeatability**

Imaging site affects repeatibility

- **UWCCC**: 18
- **MSKCC**: 10
- **NCI**: 6
- **All**: 34
Lesion location affects repeatability

<table>
<thead>
<tr>
<th>Location</th>
<th>No. lesions (%)</th>
<th>SUV\text{\textsubscript{max}}</th>
<th>SUV\text{\textsubscript{mean}}</th>
<th>SUV\text{\textsubscript{total}}</th>
<th>Coefficient of variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skull</td>
<td>9 (2.19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribs</td>
<td>78 (18.98)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervical</td>
<td>17 (4.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thoracic</td>
<td>62 (15.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lumbar</td>
<td>54 (13.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sacrum</td>
<td>31 (7.54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder</td>
<td>34 (8.27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stomach</td>
<td>15 (3.65)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arm</td>
<td>13 (3.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvis</td>
<td>70 (17.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leg</td>
<td>26 (6.81)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>411 (100%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Atlas-based segmentation identified lesion location (Yip, Perk, and Jeraj, 2014)

Repeatability poorest in ribs

CLINICAL IMPLICATIONS OF QIB IN PROSTATE CANCER

Does this make an impact?

- 56 mCRPC patients receiving multiple NaF PET/CT scans early in treatment
  - Taxane-based therapy (N = 16)
  - Androgen Receptor (AR) pathway inhibitors (N = 40)

<table>
<thead>
<tr>
<th>Baseline NaF PET/CT</th>
<th>Mid-Tx NaF PET/CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>cycle 1</td>
<td>cycle 2</td>
</tr>
<tr>
<td>cycle 3</td>
<td>cycle 1</td>
</tr>
</tbody>
</table>

- QTBI used for Quantitative NaF PET/CT analysis
Global correlation to PFS

Significant NaF PET/CT correlates

Baseline

Mid-Tx

Δ (%)

Harmon et al 2017, J Clin Oncol. 7(1):9370

Mid-Tx functional burden

High NaF burden

Low NaF burden

Harmon et al 2017, J Clin Oncol. 7(1):9370

Regional correlation to PFS

<table>
<thead>
<tr>
<th>Variable</th>
<th>HR</th>
<th>p-value (adj. p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1.41</td>
<td>0.007 (0.043)</td>
</tr>
<tr>
<td>Sacral N</td>
<td>1.71</td>
<td>0.001 (0.015)</td>
</tr>
<tr>
<td>Pelvic N</td>
<td>1.72</td>
<td>&lt;0.001 (&lt;0.001)</td>
</tr>
<tr>
<td>SUV_max</td>
<td>1.49</td>
<td>0.014 (0.053)</td>
</tr>
<tr>
<td>Lumbar SUV_max</td>
<td>1.63</td>
<td>0.004 (0.035)</td>
</tr>
<tr>
<td>Pelvic SUV_max</td>
<td>1.71</td>
<td>0.002 (0.020)</td>
</tr>
<tr>
<td>SUV_mean</td>
<td>1.70</td>
<td>0.007 (0.043)</td>
</tr>
<tr>
<td>Lumbar SUV_mean</td>
<td>1.84</td>
<td>&lt;0.001 (&lt;0.001)</td>
</tr>
<tr>
<td>SUV_total</td>
<td>1.54</td>
<td>0.002 (0.040)</td>
</tr>
<tr>
<td>Lumbar SUV_total</td>
<td>1.58</td>
<td>0.005 (0.038)</td>
</tr>
<tr>
<td>Thoracic SUV_total</td>
<td>2.04</td>
<td>0.001 (0.015)</td>
</tr>
</tbody>
</table>
Lesion-level heterogeneity

- Does heterogeneity in lesion response impact our prediction?
- 43 patients with paired baseline and mid-Tx
  - 3228 lesions tracked across scans
  - 75.2 lesions/patient (range: 3 – 315)
- Classify lesion response based on local test-retest analysis (volume dependent)
- Record the proportion of lesions contained within each response classification group


Lesion-level heterogeneity

- 40/43 patients exhibit response heterogeneity regardless of burden

Relative Lesion Burden based on $\Delta$SUV<sub>total</sub>


Relative Lesion Burden based on $\Delta$SUV<sub>total</sub>
Lesion-level heterogeneity

- 40/43 patients exhibit response heterogeneity regardless of burden
- Non-favorable response dominates progression events!

<table>
<thead>
<tr>
<th>Proportion of iSUV&lt;sub&gt;ROI&lt;/sub&gt; favorably (iCR+ iPR) responding lesions</th>
<th>Prop. of iSUV&lt;sub&gt;ROI&lt;/sub&gt; non-favorably (iPD+ iND) responding lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;31%</td>
<td>&lt;5%</td>
</tr>
</tbody>
</table>

Sample patient

- 73 year old diagnosed in 2/2012 w/ PSA=764 and diffuse bony mets (biopsy confirmed)
- Treated with combined androgen blockade; developed CRPC by 12/2012
- Palliative XRT (T6-L1) in 1/2013

Sample patient

- 73 year old, diagnosed in 2/2012 w/ PSA=764 and diffuse bony mets (biopsy confirmed)
- Treated with combined androgen blockade; developed CRPC by 12/2012
- Palliative XRT (T6-L1) in 1/2013
“Oligo-resistance”

Evolution of resistance

Medivation/Pfizer: MDV3100-18 clinical trial
All imaging aspects (coordination, analysis) managed by AIQ

Bridging the gap...

PCF Global Challenge 2014
Jeraj, Liu, Tomlins, Perlman, Simoncic
Summary

- **Quantitative Imaging Biomarkers (QIB) require:**
  - Characterization of instrument+analysis variability
    - Repeatability and Reproducibility
  - Analysis of performance characteristics
    - Limits of Agreement ⇦ Response Thresholds

- QIB enable *radically different approach* to treatment response evaluation
  - Impact on clinical decision making
  - Impact on drug development

Thanks to:

- **Image-guided Therapy Group (UW)**
  - Enrique Cuna
  - Peter Ferjancic
  - Daniel Huff
  - Christie Lin
  - Mauro Namias
  - Tim Park
  - Alison Roth
  - Matt Scarpati
  - Urban Simoncic
  - Marusa Turk
  - Damijan Valentinuzzi
  - Amy Walzman
  - Former students and postdocs...

- **Medical Physics Research Group (Slovenia)**

- **Funding**
  - PCF, NH, Medication, MV, Pfizer, UWCCC, State of Wisconsin

- **Medical Oncology/Hematology**
  - Glenn Li
  - Doug McNeil
  - Ryan Mattson
  - Mark Albertini
  - Ruth O’Regan
  - ...

- **Radiology**
  - Scott Petman
  - Tyler Brackhwar
  - Chris Jaskowiak

- **Human Oncology**
  - Paul Harari
  - Mark Vitrac

- **Medical Physics**
  - Ed Jackson

- **UWCCC TIR, CTD2, DOTs**
- **UW WONIX, NIX**