Volume Measurement at CT

Staging and Assessment of Response with Quantitative CT

Lawrence Schwartz, MD
Department of Radiology
Columbia University College of Physicians and Surgeons
LSCHWARTZ@COLUMBIA.EDU
## Recommendations

### STAGING

1. PET-CT should be used for staging in routine clinical practice and in clinical trials (category 1).

1. FDG scans can be used to image most subtypes of lymphoma and to target biopsy but is not routinely recommended in lymphomas with low FDG avidity e.g. CLL/SLL, extranodal MZL and some cutaneous lymphomas (category 1).

1. In HL and DLBCL staged by PET-CT there is no role for routine BMB. BMB is indicated only if it would change staging with a resultant change in therapy (category 1).

1. PET-CT with ceCT is desirable for staging patients likely to undergo radiotherapy ideally within a single scanning session, but a two stage approach using unenhanced PET-CT followed by regional ceCT for equivocal lesions may be preferred taking into account patient age, disease type and clinical stage (category 2).

2. Bulk remains an important prognostic factor in lymphoma. Volumetric analysis of tumour bulk and total tumour burden as well as methods combining metabolic activity and anatomical size or volumes should be explored as potential prognosticators (category 3).

3. Optimal reproducible methods for volumetric analysis are yet to be defined and will require prospective testing in multicentre studies or carefully selected retrospective datasets (category 3).
## Recommendations

<table>
<thead>
<tr>
<th>RESPONSE ASSESSMENT - QUANTITATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Standardisation of PET methods is mandatory for the use of quantitative approaches (category 1)</td>
</tr>
<tr>
<td>1. Data are emerging to suggest that quantitative measures could be used to improve on visual analysis for response assessment in DLBCL but this requires further validation in clinical trials (category 2).</td>
</tr>
<tr>
<td>1. The $\Delta$SUV$_{\text{max}}$ is the only quantitative measure with published data to indicate its possible utility in response assessment but changes in tumour volumes should also be explored (category 3).</td>
</tr>
</tbody>
</table>
Questions

- Why measure volumes?
- Are CT tumor measurements accurate?
- What type of CT acquisition is required for a volume measurement?
- How do they compare in accuracy to PET SUV measurements?
  - Do we care how they compare?
  - Combination – Metabolic Tumor Volume
- Can we perform CT volume measurements
  - At a single institution
  - In a multicenter trial
Why measure volumes?
Staging beyond Ann Arbor . . .

Max diameter 125 mm
Volume 130590 mm³

Max diameter 142 mm
Volume 215230 mm³
Why measure volumes?
Response Assessment . . .

<table>
<thead>
<tr>
<th></th>
<th>uni (mm)</th>
<th>bl (mm^2)</th>
<th>vol (mm^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>138.1</td>
<td>8446.2</td>
<td>432017.3</td>
</tr>
<tr>
<td>Follow up</td>
<td>82.4</td>
<td>3332.5</td>
<td>120840.1</td>
</tr>
</tbody>
</table>

% CHANGE  40%   59%   72%
Are CT Measurements Accurate

- 97 lymph node metastases were assessed manually (RECIST 1.1) and by volumetry with semi-automated software.

- The quality of segmentation after manual correction was acceptable to excellent in 95% of lesions and manual corrections were applied in 21 - 36% of all lesions, most predominantly in lymph nodes.

- Mean precision was 2.6 - 6.3% (manual) with 0.2 - 1.5% (effective) relative measurement deviation ($p < .001$). Inter-reader median variation coefficients ranged from 9.4 - 12.8% (manual) and 2.9 - 8.2% (volumetric) for different lesion types ($p < .001$). The limits of agreement were ± 9.8 to ± 11.2% for volumetric assessment.
Materials and Methods: MSCT scans of 63 malignant lymphoma patients before and after 2 cycles of chemotherapy (307 target lymph nodes)

Results: Response classification per lymph node revealed semi-automated volumetry and bi-dimensional WHO to be significantly more accurate than manual linear metric measurements.

Response classification per patient based on RECIST revealed more patients to be correctly classified by semi-automatic measurements, e.g. 96.0% / 92.9% (WHO bi-dimensional/volume) compared to 85.7/84.1% for manual LAD and SAD, respectively (mean reduction in misclassified patients of 9.95)
Clinically indicated non-contrast 1.25 mm slice chest CT

Repeat same CT on same scanner

Up to 15 minute break

Zhao Radiology July 2009
Reproducibility of CT Scans
Concordance Correlation Coefficient

<table>
<thead>
<tr>
<th></th>
<th>UNI</th>
<th>BI</th>
<th>VOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCC</td>
<td>0.9981</td>
<td>0.9965</td>
<td>0.9995</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.9968, 0.9994</td>
<td>0.9940, 0.9989</td>
<td>0.9991, 0.9998</td>
</tr>
</tbody>
</table>

To estimate the reproducibility and repeatability of the tumor size measurement – CCC - Used to quantify repeatability and reproducibility
Reproducibility of CT Scans
Modified Bland-Altman

Modified Bland-Altman Plot – the percentage of relative difference between the repeated tumor measurements
Reproducibility of CT Scans
2 cm example

<table>
<thead>
<tr>
<th>UNI</th>
<th>BI</th>
<th>VOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.88 cm</td>
<td>2.62 cm²</td>
<td>3.69 cm³</td>
</tr>
<tr>
<td>- 6.0%</td>
<td>- 16.6%</td>
<td>- 11.9%</td>
</tr>
<tr>
<td>2.00 cm</td>
<td>3.14 cm²</td>
<td>4.19 cm³</td>
</tr>
<tr>
<td>2.13 cm</td>
<td>3.76 cm²</td>
<td>4.47 cm³</td>
</tr>
<tr>
<td>+ 6.5%</td>
<td>+ 19.8%</td>
<td>+ 6.8%</td>
</tr>
</tbody>
</table>

For the computer generated measurements, using a hypothetical 2 cm tumor, 95% confidence interval – what would the second measurement be
What type of CT acquisition is required for a volume measurement?

Quantitative Imaging Biomarkers Alliance (QIBA)

QIBA Mission: Improve the value and practicality of quantitative biomarkers by reducing variability across devices, patients and time.

QIBA is an initiative to advance quantitative imaging and the use of imaging biomarkers in clinical trials and clinical practice by engaging researchers, healthcare professionals and industry. This involves:

- collaborating to identify needs, barriers, and solutions to develop and test consistent, reliable, valid, and achievable quantitative imaging results across imaging platforms, clinical sites, and time.
- accelerating the development and adoption of hardware and software standards needed to achieve accurate and reproducible quantitative results from imaging methods.
What type of CT acquisition is required for a volume measurement?
How do they compare in accuracy to PET SUV measurements?

How can we optimally combine the metabolic and anatomic information?
Changes in tumor metabolism are a more sensitive parameter for assessing the effect of therapy....
Metabolic tumor volume - MTV

- Volume of tumor tissues with increased FDG uptake
- FDG target volume frequently calculated by visual delineation of tumor edge or side-by-side analysis with contrast-enhanced CT scan
- Semi-automated from attenuation-corrected PET/CT images by using a contouring program, renders the volume measurement more feasible
Improving Surrogates
Automated Segmentation

Yan et al, Medical Physics 33, 2006
Can we perform CT volume measurements
Can we perform CT volume measurements
Can we perform CT volume measurements

Baseline

6-week follow-up
Can we perform CT volume measurements
Automated Segmentation
Can we perform CT volume measurements

Automated Segmentation
Can we perform CT volume measurements
Imaging and Tumor Biology
understanding response to therapy

- Used to determine *treatment decisions* for an individual patient

- Used to evaluate *efficacy of a novel therapy* in a clinical trial

- Used for *correlative analysis* to develop predictive *tissue biomarkers*
Opportunities and Questions

- Create more biologically meaningful response criteria
  - Are we using the correct cut values for PR and PD?
  - How best to evaluate the spleen

- Is the long axis or short axis a good enough surrogate for true tumor burden

- Do we need to revisit how many lesions to measure

- Is tumor burden at baseline a predictive biomarker

- Can we measure an anatomic response earlier with any of these methodologies

- Can these techniques help us define predictive tissue biomarkers
EVALUATION OF INTERIM RESPONSE IN CLASSICAL HL USING VOLUMETRIC CT MEASUREMENTS IN COMBINATION WITH FDG PET PARAMETERS AFTER 2 CYCLES

- To determine the progression-free survival (PFS) at 36 months from enrollment for patients with Hodgkin lymphoma using CT volumetric changes between baseline and after 2 cycles of AVG in combination with qualitative FDG PET/CT interpretation.

- Using changes determined by volumetric CT measurements, alone, between baseline and after 2-4 cycles and after 6 cycles of therapy and in combination with qualitative and quantitative FDG PET/CT interpretation, to determine the,
  - best overall response
  - positive and negative predictive value of each test metrics alone and in combination with each other.
  - compare the predictive values of combinatorial imaging (vCT and FDG PET/CT) parameters with conventional risk factors including IPI.
Answers...

- Why measure volumes?
- Are CT tumor measurements accurate?
- What type of CT acquisition is required for a volume measurement?
- How do they compare in accuracy to PET SUV measurements?
- Do we care how they compare?
- Can we perform CT volume measurements
  - At a single institution
  - In a multicenter trial