

Development of an Expert Consensus Target Delineation Atlas for Thymic Cancers: Initial Quantitative Analysis of an Expert Survey.

Charles R. Thomas, Jr., MD, Jayashree Kalpathy-Cramer, PhD,
Jehee Choi, BS, Daniel Gomez, MD, Andreas Rimner, MD, Ying Li,
MD, PhD, Suresh Senan, MD, PhD, Lynn Wilson, MD, MPH,
Ritsuko Komaki, MD, Clifton D. Fuller, MD, PhD



ITMIG

International
Thymic
Malignancy
Interest Group

Introduction

- Conventional radiotherapy for thymic neoplasms results in potentially preventable sequelae
- Earlier reports found grade 3 and 4 toxicity ranging from 11 to 13%¹
- However, newer reports suggest a decreased rate of 5 to 10%, potentially due to technical innovations in RT delivery.²

1-Hsu HC, Huang EY, Wang CJ, et al. Postoperative radiotherapy in thymic carcinoma: treatment results and prognostic factors. *IJROBP* 2002;52:801– 805

2-Lucchi M, Mussi A, Basolo F, et al. The multimodality treatment of thymic carcinoma. *Eur J Cardiothorac Surg* 2001;19:566–569

ITMIG

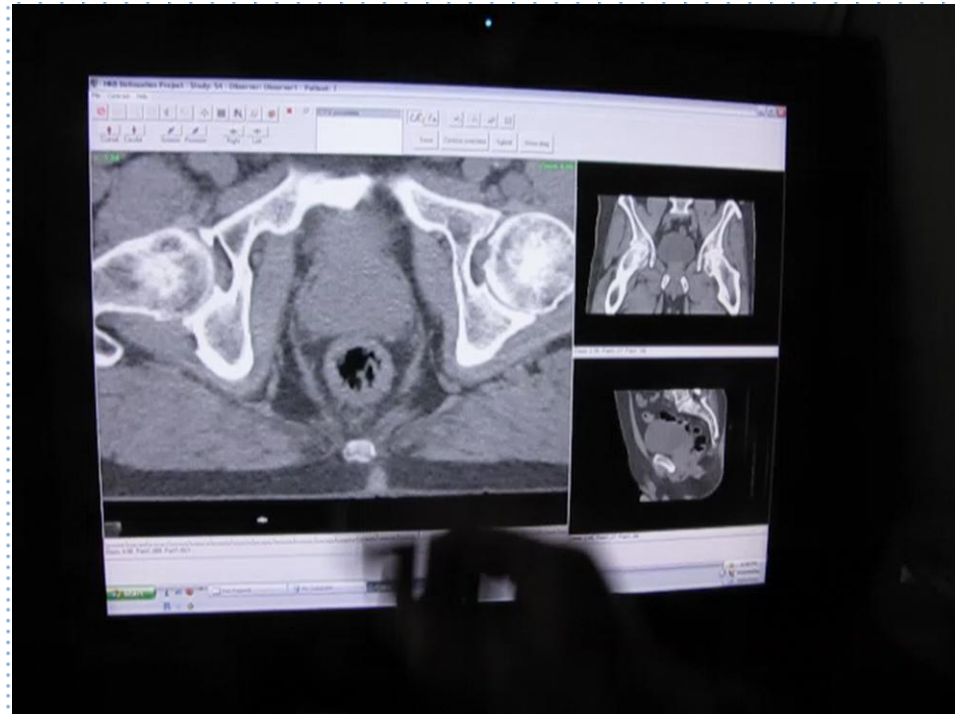
International
Thymic
Malignancy
Interest Group

Introduction II

- The advent of conformal radiotherapy requires accurate target delineation, or risks:
 - geometric misses of target volumes (TV) >> tumor recurrence
 - Over-treatment of adjacent organs-at-risk (OAR)>> excessive preventable sequelae

Target delineation

- Crucial to the RT process
- Time consuming
 - ~1-2 hours per case
- Operator dependent



Target Delineation

- GIGO=garbage in/garbage out

therapy and even more with IMRT. It has been stated that inadequate definition of the GTV/CTV leads to a systematic geographic miss of the tumor and that, for some locations, these inconsistencies may dominate all other errors in radiotherapy planning and delivery [31]. To help understand, at

The Reasons for Discrepancies in Target Volume Delineation

Wendy Jeanneret-Sozzi¹, Raphaël Moeckli², Jean-François Valley², Abderrahim Zouhair¹, Esat Mahmut Ozsahin¹, René-Olivier Mirimanoff¹ on Behalf of SASRO³
Strahlenther Onkol 2006 · No. 8

ITMIG

International
Thymic
Malignancy
Interest Group

Target delineation for thoracic malignancy

- Substantial variability in target delineation
- Worse when tumor abuts mediastinum/pleura as opposed to parenchyma
- As most thymic malignancies arise in the anterior mediastinum, most will abut OARs (heart) raising the risks for “over-contouring” tumor

Fitton et al.

- The magnitude of the interobserver SD is ~ 8 mm with CT alone for mediastinal NSCLC

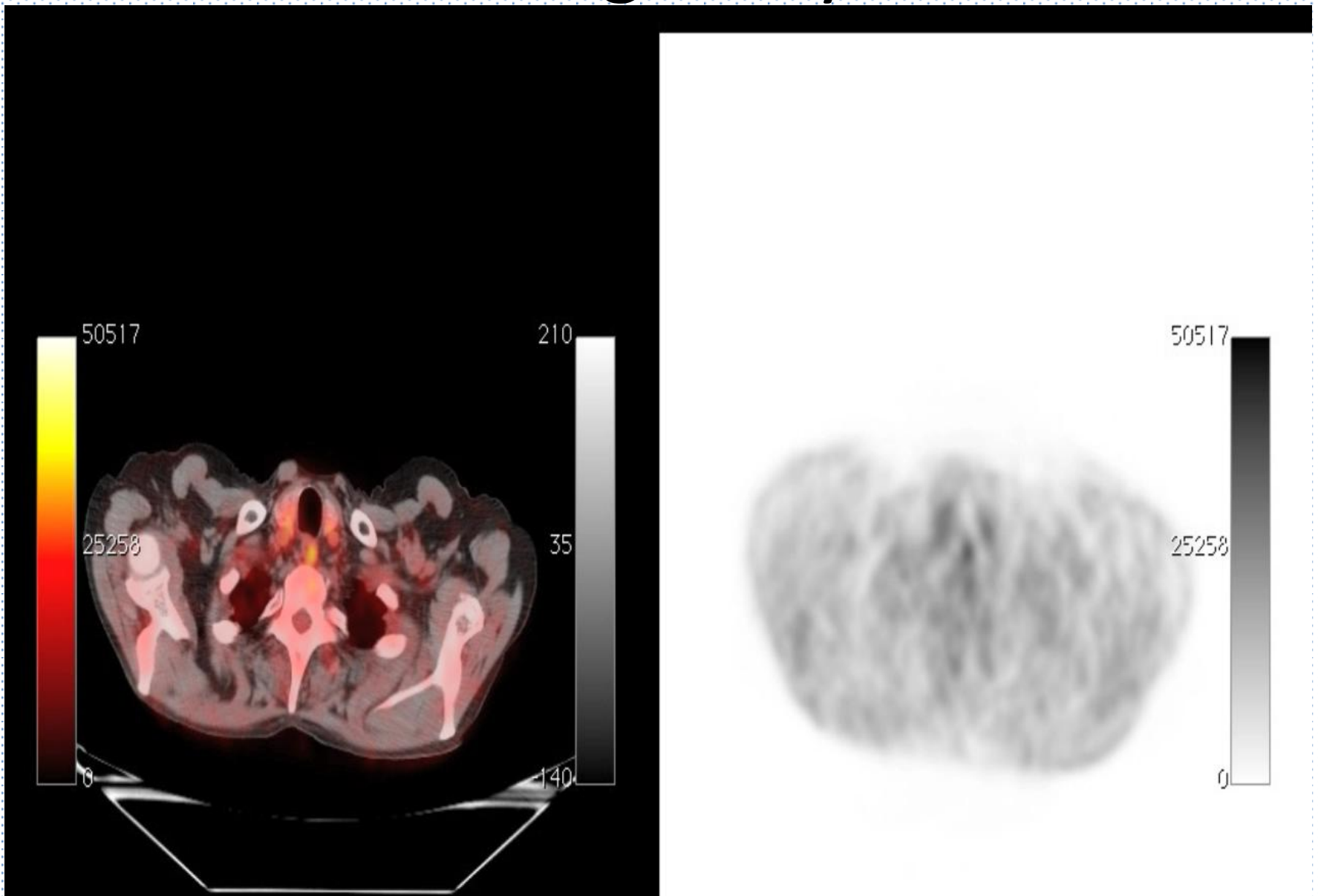
Table 2. Interobserver variability for primary tumors in Group II (*i.e.*, invading the hilar region, heart, large vessels, pericardium, or mediastinum over more than one quarter of its surface)

Group II SD (cm)	CT	CT+PET
Lung/tumor	0.8 ± 0.2	0.3 ± 0.1
Mediastinum, hilum/tumor	0.8 ± 0.2	0.4 ± 0.1
Atelectasis/tumor	1.3 ± 0.5	0.5 ± 0.1
Total	1.0 ± 0.3	0.4 ± 0.1

Abbreviations: CT = computed tomography; PET = positron emission tomography; SD = standard deviation.

Results are given for several interfaces on CT vs. CT/PET scans. In this group, the gain from using PET is very great.

Target delineation for thoracic malignancy



Target delineation for thoracic malignancy

- Variability of this magnitude BEFORE accounting for set-up error is potentially meaningful in the context of cooperative studies using IMRT/IGRT
- This provides a rationale for atlas based interventions

Atlas interventions

- Shown to be effective in GI, breast, and head and neck sites
- However, limited improvement in lung delineation seen despite an elaborate educational intervention by Vorwerk et al.:

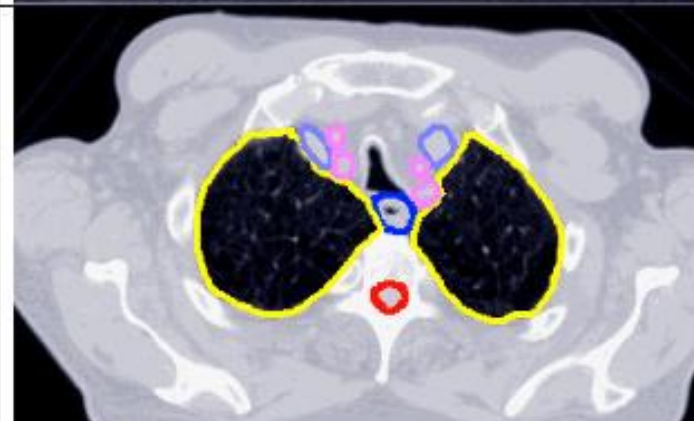
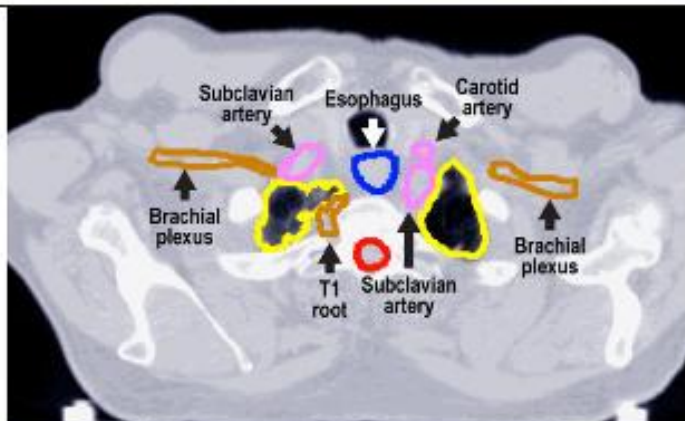
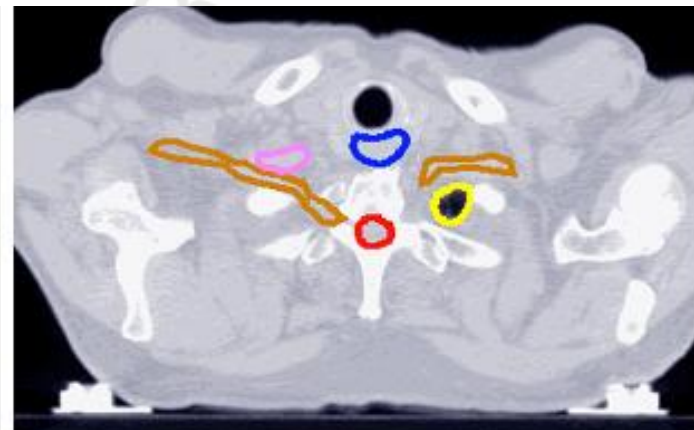
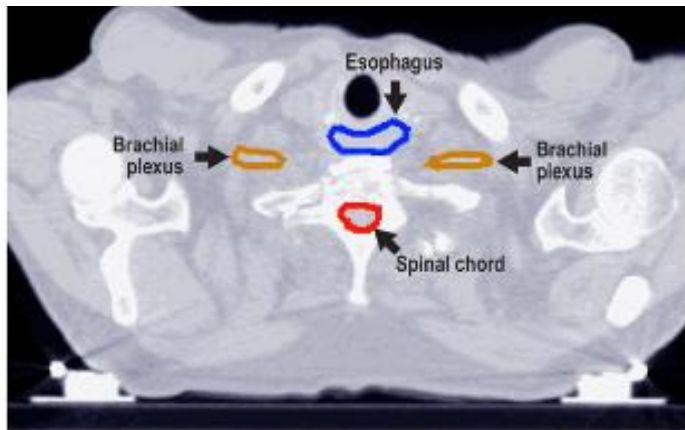
Nevertheless, we could not demonstrate an improvement in agreement between the physicians by use of a detailed instruction for delineation compared to existing literature with no detailed instructions, but we could demonstrate, that physicians from one department agree more than physicians from different departments. We recommend online available examples and the possibility of comparison of the own delineation in order to control oneself periodically for better agreement in contouring of target volumes.

Vorwerk et al.
Radiother Oncol. 2009
Jun;91(3):455-60

Atlas-based interventions

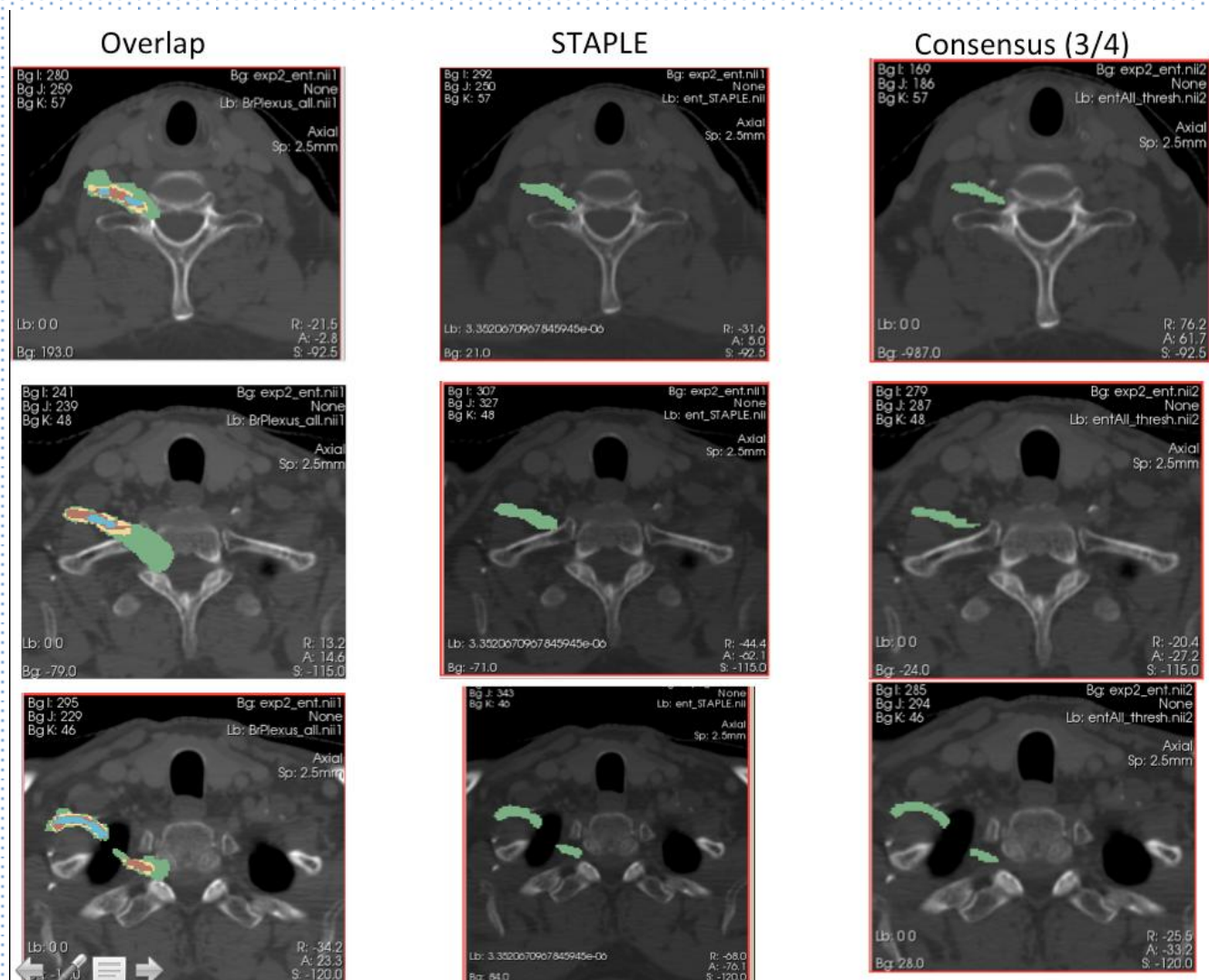
CONSIDERATION OF DOSE LIMITS FOR ORGANS AT RISK OF THORACIC RADIOTHERAPY: ATLAS FOR LUNG, PROXIMAL BRONCHIAL TREE, ESOPHAGUS, SPINAL CORD, RIBS, AND BRACHIAL PLEXUS

FENG-MING (SPRING) KONG, M.D., PH.D.,* TIMOTHY RITTER, PH.D.,* DOUGLAS J. QUINT, M.D.,[†]
SURESH SENAN, M.D.,[‡] LAURIE E. GASPAR, M.D.,[§] RITSUKO U. KOMAKI, M.D.,[¶]
COEN W. HURKMANS, PH.D.,^{||} ROBERT TIMMERMAN, M.D.,[#] ANDREA BEZJAK, M.D.,**
JEFFREY D. BRADLEY, M.D.,^{††} BENJAMIN MOVSAS, M.D.,^{‡‡} LON MARSH, C.M.D.,* PAUL OKUNIEFF, M.D.,^{§§}
HAK CHOY, M.D.,[#] AND WALTER CURRAN, M.D.,^{¶¶}



A pilot study evaluating TaCTICS software as a tool for QA in stereotactic radiotherapy of lung tumors in close proximity to critical structures

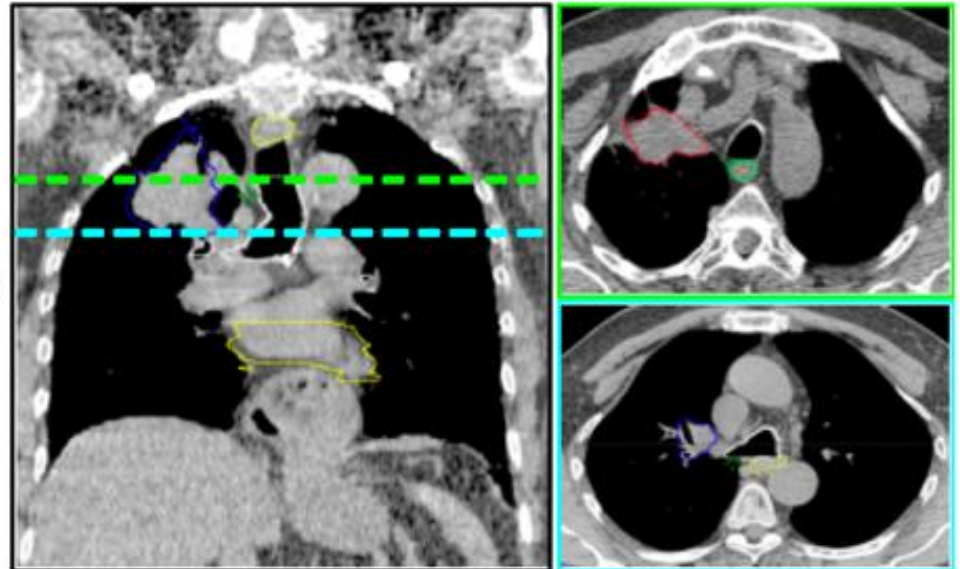
Bongers et al.



Bongers et al.

C. Results

The level of agreement between each resident's contour and the reference contour was highest for tumor and heart, with Dice similarity coefficients ranging between 0.83-0.86 and 0.87-0.92, respectively. Lower levels of agreement were observed for esophagus and proximal bronchial tree with similarity coefficients between 0.51-0.72 and 0.58-0.66. The image is an example of what is provided to the user and illustrates differences between one of the resident's contours and the reference.



So how do we make target delineation better in thymic neoplasms?

Purpose

- We sought to quantitatively determine the relative interobserver variability of expert target volume delineation as part of a larger standardization effort to develop an ITMIG consensus contouring atlas to complement radiotherapy recommendations for thymic cancers.

Specific aims

- The overall specific aims of this ITMIG-supported effort include:
 - Quantification of benchmark expert interobserver variability in target delineation*
 - Development of an expert consensus atlas for radiotherapy target delineation of thymic tumors
 - Assessment of the impact of multimodality imaging inputs

*-current report

Hypothesis

- Expert target delineations for thymic cancers would demonstrate levels of variability comparable to NSCLC contouring analyses
 - (e.g. mean Dice similarity coefficient [DSC] between 0.6-0.8).

Case posting for consensus phase

- Currently, we plan for 4 cases to be contoured
- Multimodality cases (CT only, CT-PET, 4DCT, and CT-MR datasets)
- Definitive and adjuvant cases
- Expected contouring time ~1.5 hr/case
- This report is the data collected from Case 1 (CT only)



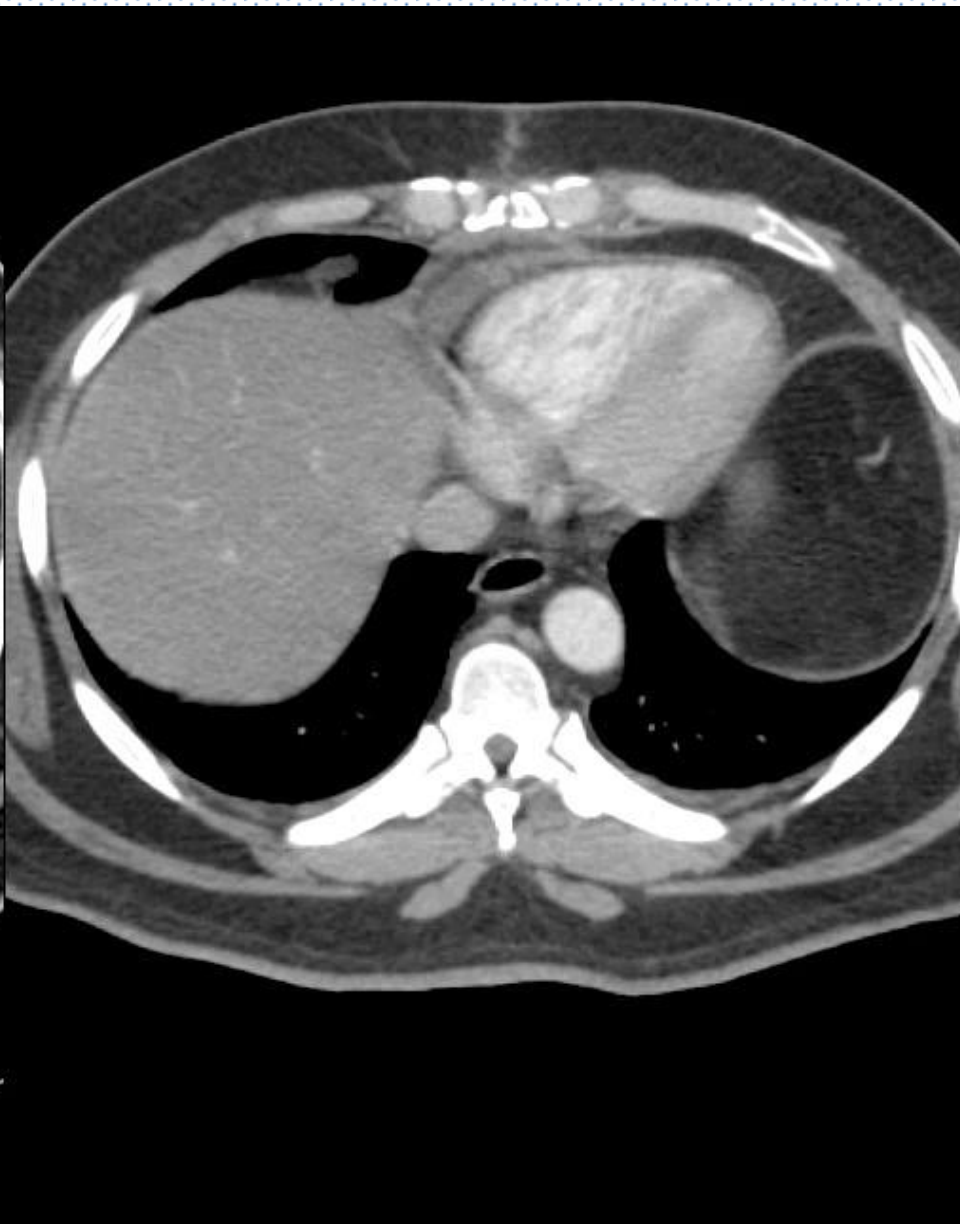
Methods I

- A pilot dataset was made of a standardized case presentation with anonymized pre- and post-operative DICOM CT image sets from a single patient with Masaoka-Koga Stage III thymoma
- These were posted for download using the TaCTICS (Target Contour Testing/Instructional Computer Software) web interface
- (http://skynet.ohsu.edu/tactics_ITMIG).

Pre-op



Post-op



Methods III

- Participating radiation oncologists contoured, using a treatment planning system of their choice, gross tumor volume (GTV), and clinical target volumes (CTV)
- A definitive (pre-op scan) and adjuvant (post-op) case were presented
- DICOM RTSTRUCT files were sent for central analysis with TaCTICS software
- users completed a survey detailing the dose prescription and PTV margins they would recommend in definitive and post-operative (i.e. R1 vs R2) scenarios.

Methods

- Interobserver variability was analyzed quantitatively with Warfield's simultaneous truth and performance level estimation (STAPLE) algorithm to generate a composite segmentation estimate of a "ground truth" target volume,
- This was then compared to each individual experts contours (Figure 2) as a Dice similarity coefficient (DSC) [where $DSC=1$ indicates total agreement, while $DSC=0$ indicates no overlapping TV voxels] using nonparametric analysis.

Consensus atlas generation

- TaCTICS currently has capacity for consensus ROI generation from multiple users using the following methods
 - Overlap of all users
 - Probabilistic overlap (e.g. X% of users contouring a voxel)
 - Warfield's STAPLE algorithm, a maximum likelihood estimation

Results I

- Seven users completed the contouring tasks for definitive and post-operative cases; of these 5 completed online surveys.
- Segmentation performance was assessed, with high mean \pm SD STAPLE-estimated segmentation sensitivity for:
 - definitive case GTV and CTV at 0.77 and 0.80, respectively
 - and post-operative CTV sensitivity of 0.55;
 - all volumes had specificity of ≥ 0.99 .

Results II

- Inter-observer agreement was markedly higher for the definitive case target volumes, with mean \pm SD DSC of 0.88 ± 0.03 and 0.89 ± 0.04 for GTV and CTV respectively, compared to post-op CTV DSC of 0.69 ± 0.06 (Kruskal-Wallis $p < 0.01$; Figure 1) for all experts vs. the STAPLE composite(s).

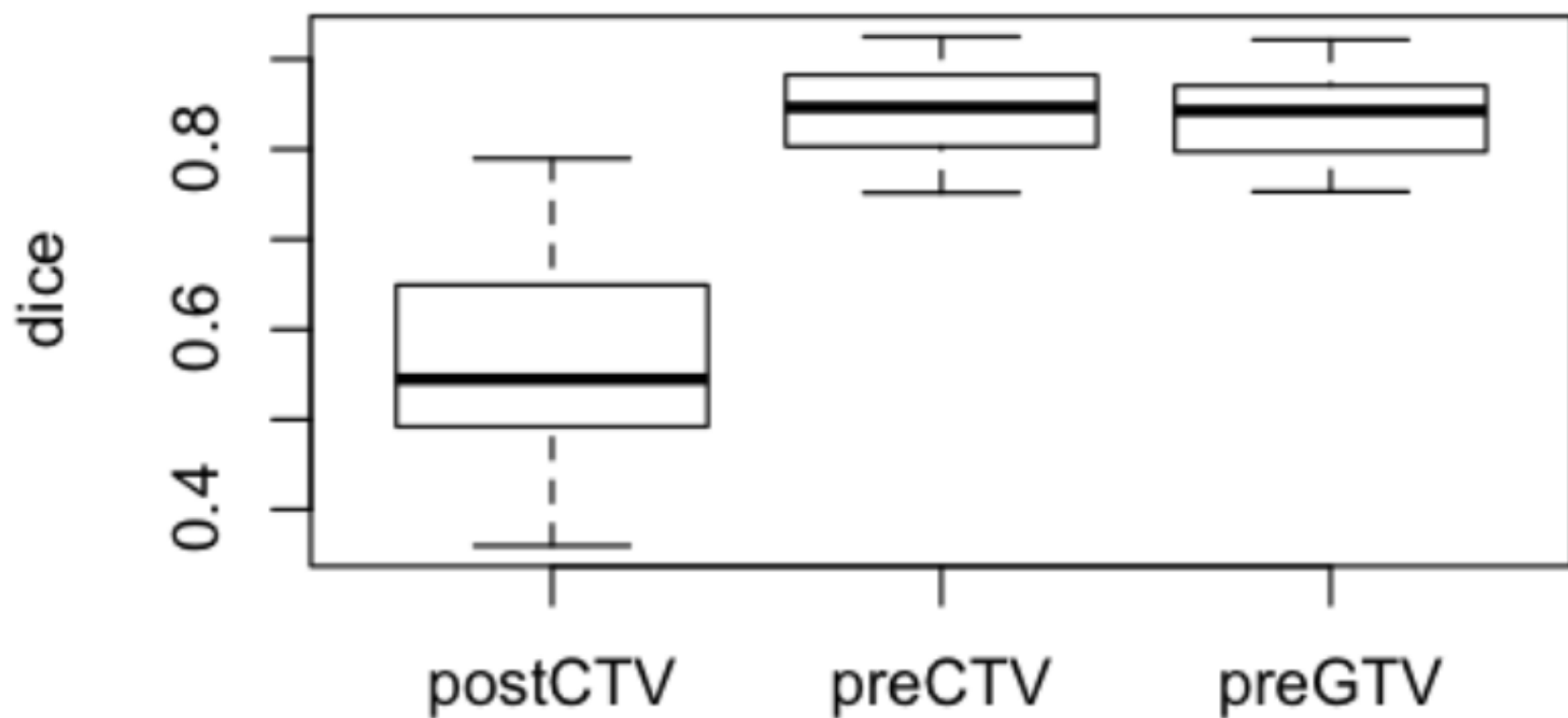


Figure 1: Dice similarity coefficients distribution of all expert participants, as compared to STAPLE-estimated ground truth (postCTV=post-operative CTV; preGTV=definitive case GTV; preCTV=definitive case CTV), where DSC=1 indicates total volumetric agreement.

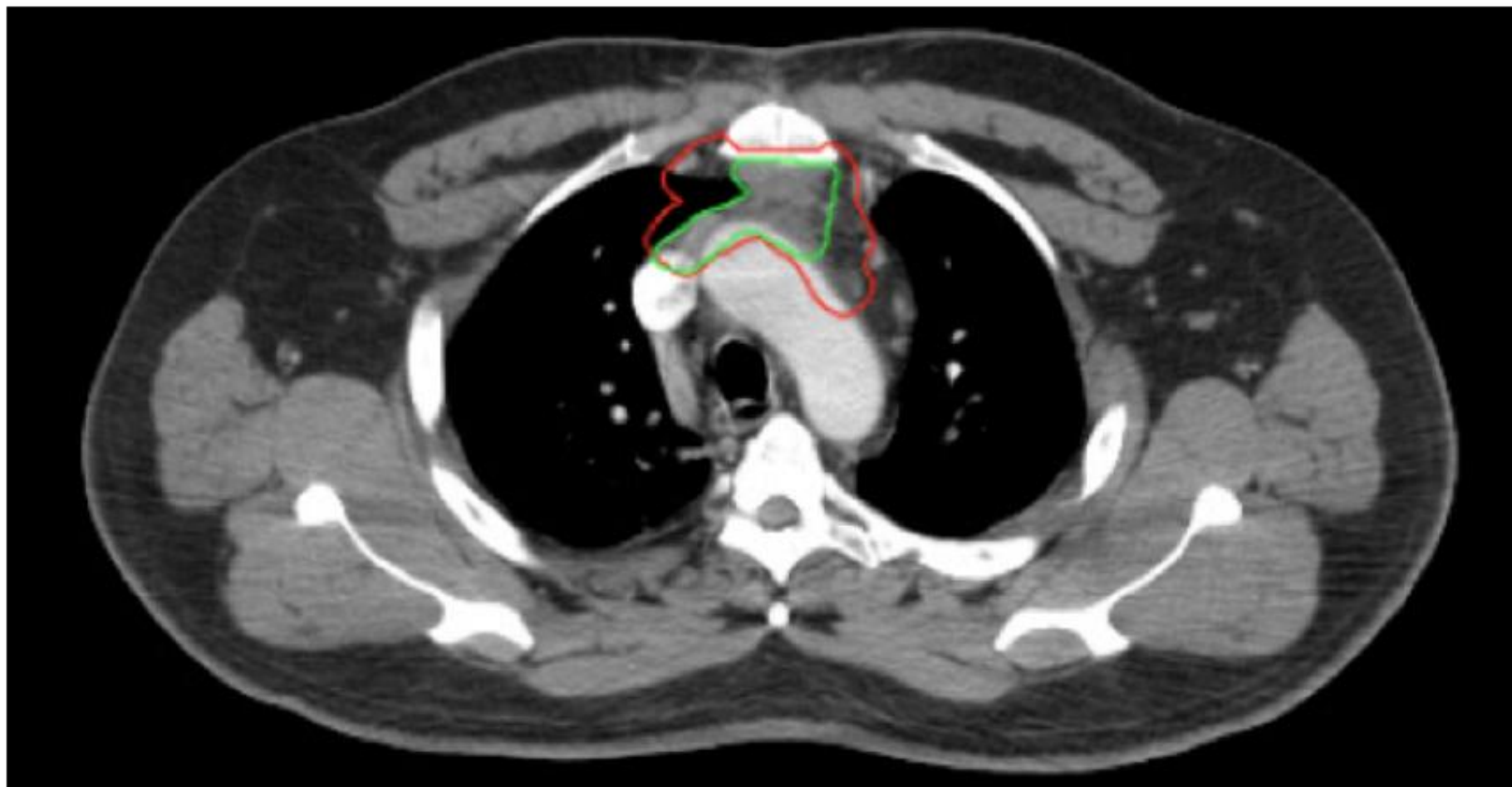


Figure 2: Screenshot of TaCTICS software, simultaneously displaying two expert-segmented target volumes (green and red outlines).

Survey results

- Survey results indicated 60% of experts would routinely add margins of 0.5 mm to CTV volumes for PTV generation, with 40% advocating 1 cm expansions.
- All respondents suggested IMRT be used, with 80% suggesting 4DCT, and 40% PETCT implementation.
- For definitive cases, a PTV prescription of 50-54 Gy was suggested.
- Post-operatively, for R1 resections, a PTV prescription of 60-66 Gy was suggested, with 60-70 Gy for an R2 resection

Conclusions

- Expert agreement for definitive case volumes was exceptionally high, markedly better than comparable thoracic cases.
- However, significantly lower agreement was noted post-operatively.
- Technique and dose prescription between experts was substantively consistent (which is reassuring).

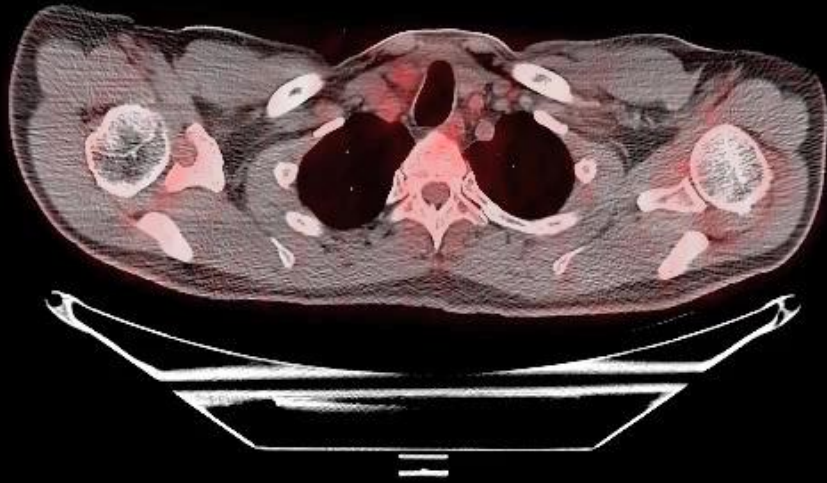
Future directions

- 3-4 more cases planned for atlas development phase.
- Data from this preliminary analysis will be used to generate educational materials to assist non-expert target delineation through guideline and atlas implementation.
- Future efforts will seek to assess the impact of MRI-CT, 4DCT and PET-CT imaging on thymic target delineation variability.

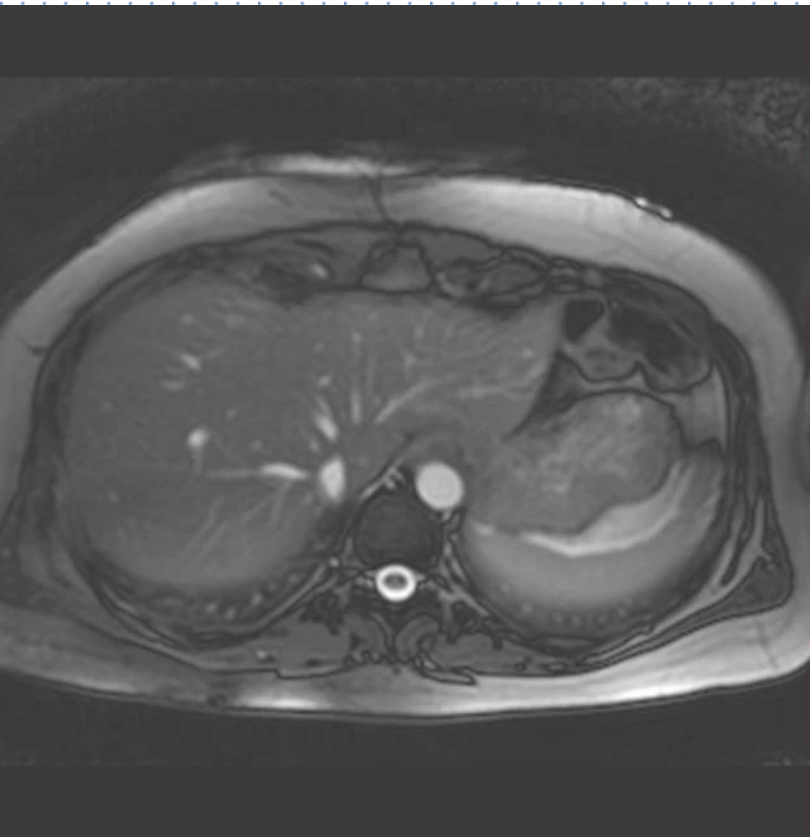
Multimodality image usage

- Is this:

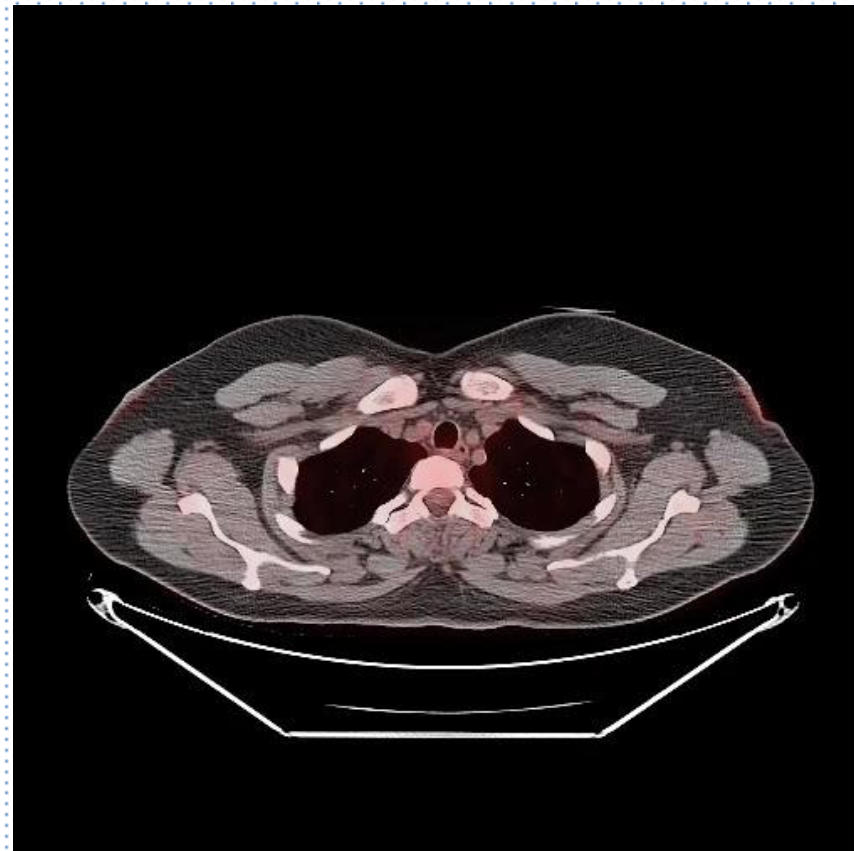
any better than this:



How do we use advanced MR data?



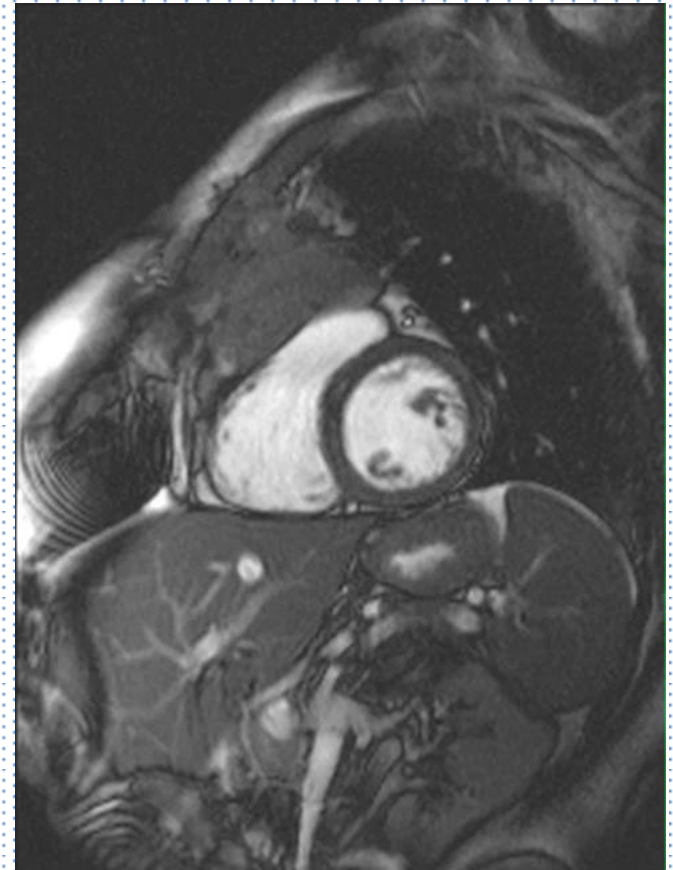
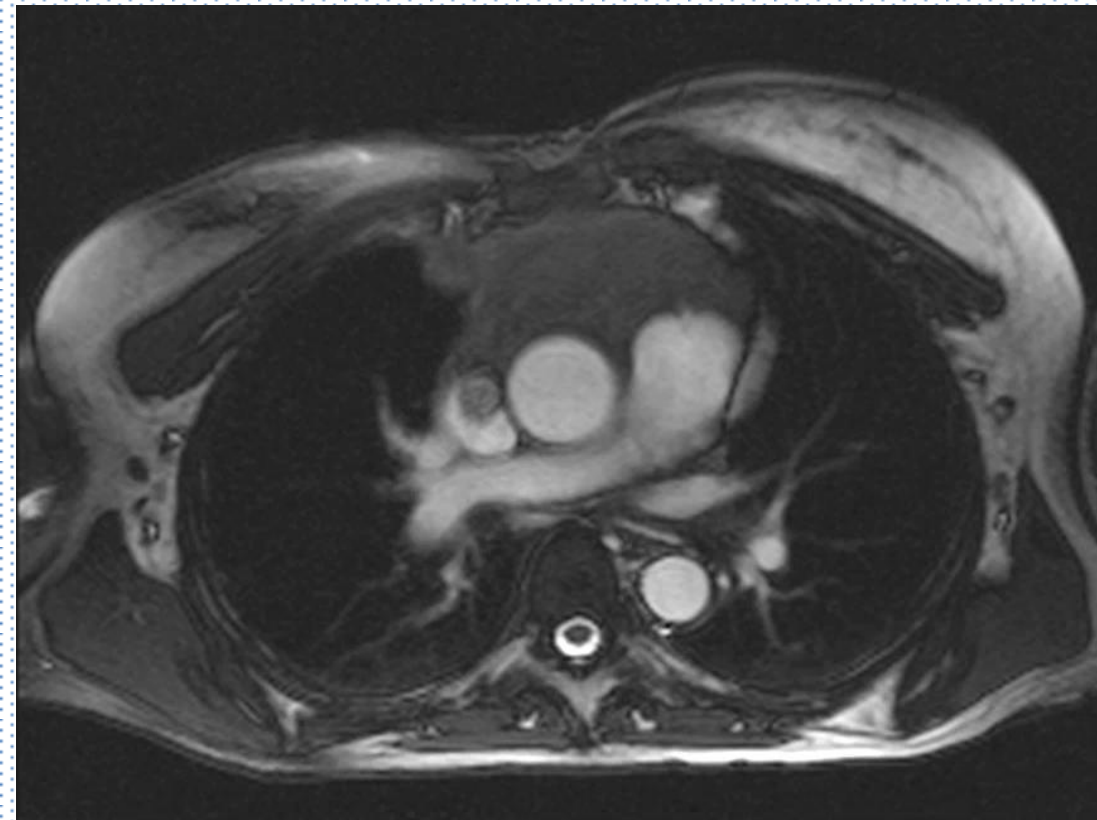
+



ITMIG

International
Thymic
Malignancy
Interest Group

What about motion correction?



ITMIG

International
Thymic
Malignancy
Interest Group

Next

- We are preparing to release the next cases for the expert contouring Phase.
- We will then use expert contours to create an online atlas/manuscript.
- If anyone is interested in participating, please email me at thomasch@ohsu.edu

THANKS!