

## AbstractID: 12784 Title: Validation Study of a Software Tool for Consensus Analysis of Experts' Contours for Generating Atlases of Radiotherapy Target and Normal Structures

**Purpose:** To analyze and validate the ability of a newly developed software tool to estimate and analyze consensus contours from manually delineated structures by panel of expert radiation oncologists for building reference atlases for radiotherapy and dosimetry practitioners. **Method and Materials:** Several statistical methods and a graphical user interface were developed as an integral part of an in-house research treatment planning system. For evaluation purposes, we used three breast cancer CT scans from the RTOG Breast Cancer Atlas Project labeled A, B, and C. All participating physicians contoured specific structures before and after the consensus panel met without the software aid. Differences in the contours were analyzed qualitatively and quantitatively using metrics based on apparent agreement, kappa statistics, and STAPLE estimates by the developed consensus software tool. Estimates of the consensus contours were analyzed for the different structures before consensus and compared with the experts' manual consensus contour. Dice similarity coefficient (DSC) was used for comparative evaluation between the software and the experts' results. **Results:** Dice similarity calculations for all pre-consensus STAPLE estimates and the final consensus panel structures reached 0.80 or greater for the heart, left breast, right breast, case A lumpectomy, case B lumpectomy, and the chest wall. The heart, case A lumpectomy, left breast, and right breast all exhibited a DSC greater than 0.90. The highest Dice coefficients were seen in the comparison of the STAPLE estimated left breast contours, which almost reached 0.96 at 100% confidence level. DSCs were in the range of .70 to .79 for the supraclavicular and internal mammary nodes. **Conclusion:** Using the consensus software tool incorporating several statistical methods allowed excellent ability to create contours similar to the ones generated by expert physicians and provided metrics for the confidence levels of agreement.