Purpose/Objective:
Statistically trainable deformable-shape models (SDSMs) of anatomic objects show promise for segmentation by automatically deforming in a target image to closely match target anatomy. The purpose was to establish benchmarks for comparison with clinical studies for a particular class of SDSMs called m-reps in segmenting the prostate and bladder for IGRT/ART.

Materials/Methods:
79 CT treatment images for 5 patients undergoing IGRT for prostate cancer were manually contoured by an expert. M-rep models were fit to the contours and statistically analyzed to compute a mean for each patient and a probability distribution of day-to-day within-patient position and shape variability of the bladder and prostate. Each fitted m-rep also was registered with its corresponding image to collect, in m-rep coordinates, histograms of image intensities in predefined local regions inside and around the prostate and bladder. The histograms were processed to yield probability distributions on within-patient anatomy-relative image intensity patterns. In a leave-one-patient-out experiment, mean m-rep models were applied to each treatment image for each set of patient images, where the target patient's images were excluded from training data. The mean-to-date prostate m-rep for the target patient was initialized in each target image via an automatically computed transformation matrix that registered the target image with the first treatment image. Due to its wide shape variability the mean-to-date bladder m-rep was initialized via three manually drawn contours near the base, middle and superior sections. The deformations were automatically driven by an algorithm that optimized an objective function with terms for the geometric and intensity probability distributions described above. These studies included global deformations for the prostate, and global and regional deformations for the bladder.

Results:
For the prostate the median volume overlap between m-reps and the human expert was 94% for both the prostate and bladder. The median of the average surface separation between m-rep and human segmentations over all cases was 0.9 mm for the prostate, and 1.0 mm for the bladder. This level of agreement is at voxel scale and is better than two human experts would agree.

Conclusions:
M-rep optimization mimics contouring by a human expert against which m-reps have been trained. Moreover, assuming intra-observer variabilities are similar across experts, special properties of m-reps allow the position and shape variabilities to apply to a mean model estimated by another expert, e.g., by manual contouring the planning image, without additional training. To encourage open comparison and publication of segmentation results, images and associated data sets used in this study are available to other investigators. This work was supported by NIH/NIBIB grant P01 EB002779.