Analysis of Whole-Body CT Segmentation Models for Pediatric Organ Dosimetry in Clinical Workflows

Elanchezhian Somasundaram PhD, Bryan Luna BS, Will Tepe BS, Joe Meier PhD, Alex J Towbin MD, Samuel L Brady PhD
Cincinnati Children's Hospital Medical Center

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

INTRODUCTION

Background:

- Automatic organ segmentation is essential for patient-specific CT dosimetry.
- While AI-based whole-body segmentation tools are routinely used for adult Organs at Risk (OARs), the high variability in pediatric anatomy requires validating these models on real-world pediatric data.

Objective:

- To evaluate the reliability of a benchmark AI segmentation tool across ICRP OAR categories and pediatric age groups to determine its readiness for clinical pediatric dosimetry workflows.

METHODS

Model Deployment & Data Extraction

- The MONAI CT whole-body segmentation (MONAI_WB) model - a SegResNet-based version of the original CT TotalSegmentator was deployed at a large pediatric hospital via MONAI Deploy with automated routing of clinical CT exams over a 4-month interval.
- The latest nnUNet-based CT TotalSegmentator (Latest_TS) was deployed and retrospectively run on the same cohort for paired comparison of the model versions and their differences.

Stratification & Grouping

- The study dataset consisted of 2193 paired exams segmented by both model versions.
- Zero-volume organs were treated as outside the imaged field of view and excluded.
- Comparison was restricted to 78 shared component organs mapped into 9 OAR groups based on ICRP 103 tissue-weighting categories, including brain, esophagus, liver, bladder, lungs, colon, stomach, red marrow surrogates, and remainder tissues.
- Component organ volumes were summed within each OAR group for each exam.
- Age groups were defined as: 0–2y, 2–5y, 5–9y, 9–14y, 14–18y, and ≥18y.

Analysis

- Statistical summaries were performed within age and OAR region groups.
- **Within-model variability:** Evaluated using CoV (Coefficient of Variation).
- **Outlier burden:** Quantified using the IQR (Interquartile Range) rule.
- **Cross-version agreement:** Assessed using paired volume-difference and relative-difference metrics, including CoV difference (CoV_{Δ}).

CONTACT

Elanchezhian Somasundaram (Elan)
Cincinnati Children's Hospital Medical Center
3333 Burnett Avenue, Cincinnati, Ohio, USA 45229
Elanchezhian.Somasundaram@cchmc.org
513-636-3164

RESULTS

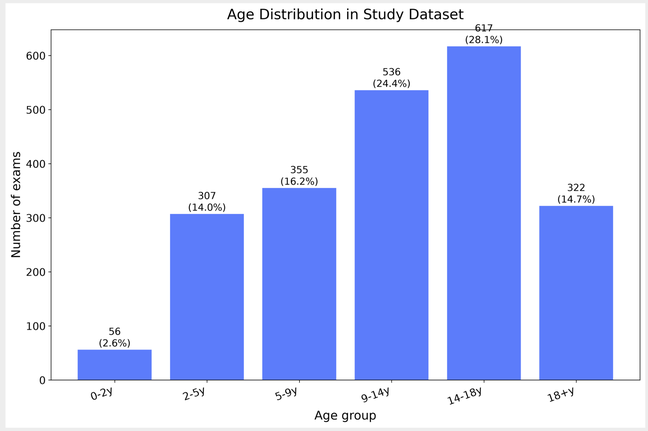

Figure 1. Age Distribution.

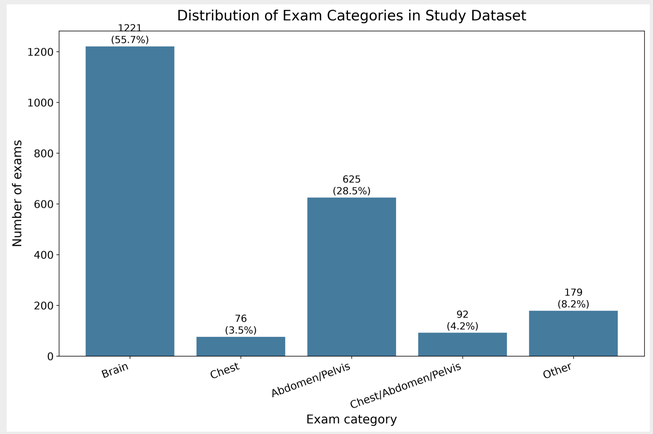

Figure 2. Study Distribution

Coefficient of Variation and Outliers by OAR Groups:

- For Latest_TS model, red marrow surrogates showed the greatest variability (mean CoV 1.34), followed by bladder (0.82), esophagus (0.75), and lungs (0.75), whereas liver (0.33) and brain (0.34) were the least variable OAR groups.
- CoV - Age dependence was OAR-specific rather than uniform: red marrow surrogates were most variable in younger patients and became progressively more stable with age (CoV 1.54 at 2-5 years vs 1.10 at 18+ years), remainder tissues showed a similar decline (0.86 at 0-2 years vs 0.43 at 18+ years), and brain maintained low CoV across ages (0.25-0.39).
- Outlier burden for Latest_TS model was highest for brain (7.3%), esophagus (6.6%), and red marrow surrogates (5.1%), and lowest for remainder tissues (0.8%) and colon (2.0%).
- Brain concentrated its outlier burden in children 0-14 years (8.2-12.9%) with minimal outliers after age 14.

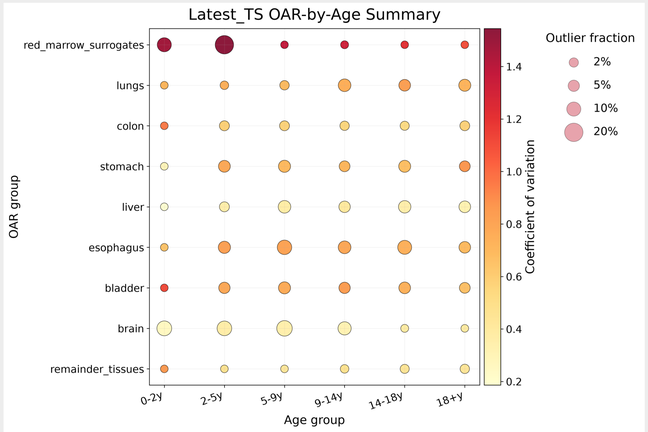

Figure 3. Age-group wise CoV and Outliers in OAR volume for Latest_TS model

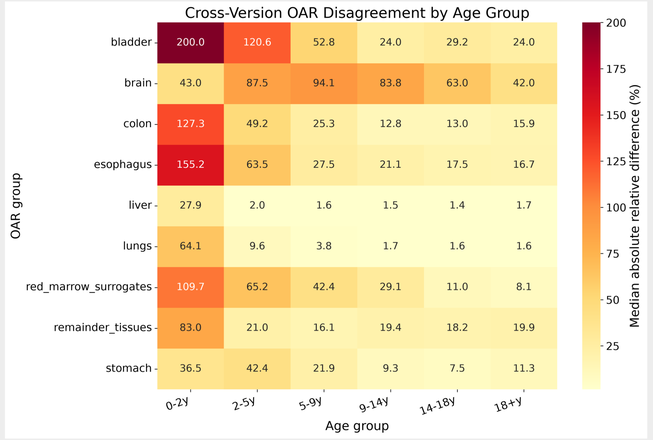

Figure 4. Model Cross-version OAR disagreement by age group

Model version disagreement:

- Relative to MONAI_WB model, the largest cross-version disagreement was seen in brain (median absolute relative difference 73.0%), followed by bladder (35.3%), red marrow surrogates (24.8%), and esophagus (21.5%); liver (1.59%) and lungs (2.26%) were the most stable OAR groups.
- Variability shifts between versions were non-uniform: brain showed the largest reduction in CoV from MONAI_WB to Latest_TS (0.733 to 0.358; CoV_{Δ} -0.375), whereas stomach showed the largest increase (0.694 to 0.822; CoV_{Δ} +0.129).
- A strong pediatric age effect was present between model versions: bladder disagreement peaked in the youngest groups at 200.0% for ages 0-2 years and 120.6% for ages 2-5 years, while brain disagreement dominated older groups at 94.1% for ages 5-9 years and 83.8% for ages 9-14 years.

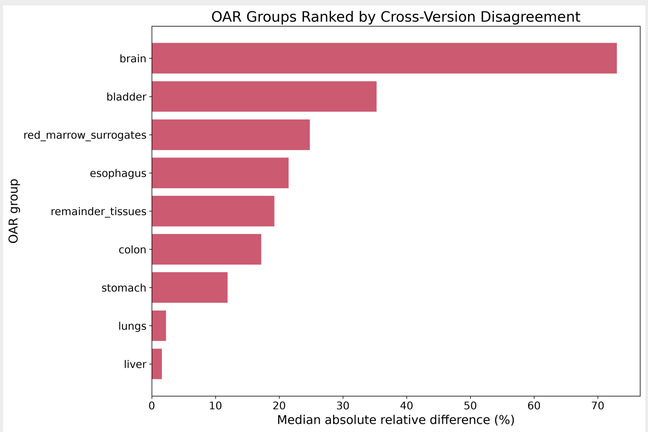

Figure 5. Overall disagreement in OAR relative volumes across model versions

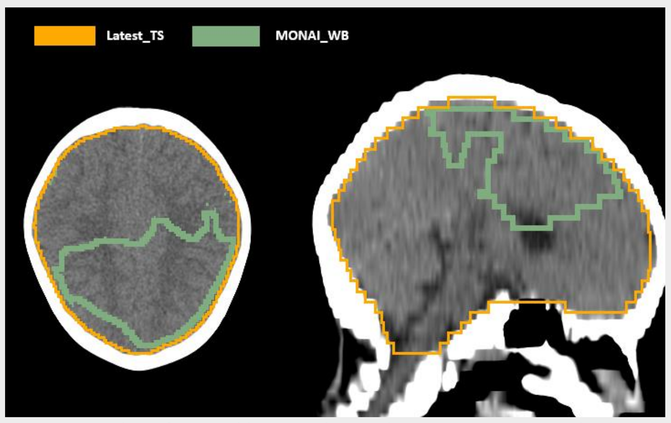

Figure 6. Head CT of a 6-yr-old patient showing brain mis-segmentation by the MONAI_WB model, compared to accurate segmentation by the Latest_TS model

DISCUSSION

- **CoV-based variability:** Within Latest_TS, red marrow surrogates showed the greatest exam-to-exam volume variability, with bladder, esophagus, and lungs also relatively variable, whereas liver and brain had the lowest overall CoV.
- **Outlier-based variability:** Brain, esophagus, and red marrow surrogates had the highest fractions of extreme observations, while remainder tissues and colon had the lowest outlier burden; notably, brain paired low average CoV with the highest outlier fraction, suggesting mostly consistent segmentations with occasional discrete failures.
- **Cross-version disagreement:** Brain, bladder, red marrow surrogates, and esophagus showed the largest shifts between MONAI_TS and Latest_TS, whereas liver and lungs were the most stable shared OAR groups, indicating that low within-model variability in Latest_TS does not necessarily imply cross-version agreement.
- **Age dependence**: Red marrow surrogates and remainder tissues showed the strongest age dependence in Latest_TS CoV, with higher variability in younger children and greater stability with age; brain CoV remained low across age groups, but pediatric outlier burden was concentrated in ages 0-14 years.
- **Age dependence in disagreement:** Bladder and red marrow surrogates showed the strongest age dependence in model-to-model disagreement, with the largest differences in the youngest children, whereas brain disagreement remained elevated across multiple age bands.

CONCLUSIONS

- **Segmentation Reliability:** AI segmentation performance is heavily dependent on the specific organ, metric, and model version. While liver and lungs showed stability, structures like the brain and red marrow exhibited high outlier burdens and significant cross-version disagreement.
- **Age-Related Variance:** Pediatric age heavily influences model performance, with the youngest cohorts displaying the highest variability (CoV) in red marrow and remainder tissues, and the highest outlier burden in brain segmentations.
- **The Need for Pediatric Models:** Relying on adult-benchmarked segmentation models is insufficient for clinical dosimetry in children. Developing dedicated, pediatric-specific whole-body composition models is essential to ensure accurate risk estimation across all ages and OAR groups.

REFERENCES

1. ICRP Publication 103. The 2007 Recommendations of the International Commission on Radiological Protection. Ann ICRP 2007;37(2–4).
2. MONAI: Medical Open Network for AI. Project-MONAI/MONAI. monai.io.
3. Cardoso MJ, et al. MONAI: An open-source framework for deep learning in healthcare. arXiv:2211.02701, 2022.
4. Wasserthal J, et al. TotalSegmentator: robust segmentation of 104 anatomical structures in CT images. arXiv:2208.05868, 2022.