



QC-assisted novice acquisition yields comparable performance to trained acquisition without QC during obstetric sweep
antenatal ultrasound screening

Sean Flannery, Elizabeth Herbst, Priyam Patel, Manikanda Krishnan, Stephen Schmidt, Emily Cai, Melanie Jutras, Madhumita Gupta, Anuradha Rao, Rono Kiplimo, Shyam Bharat, Jonathan Sutton
Philips Ultrasound (Cambridge, MA, USA; Bangalore, IN; Nairobi, KE)

Background

- AI-enabled obstetric sweep (OS) workflows could expand antenatal risk stratification when traditional ultrasound screening is unavailable
- During an OS exam, transabdominal images are acquired in a grid pattern and interpreted by AI
- Transducer angle and coupling degrade gestational age (GA) and amniotic fluid deepest vertical pocket (DVP) results
- Providing quality control (QC) feedback to novice users to maintain correct angle and coupling may improve clinical results for GA and DVP

Hypothesis

QC-assisted novice users will achieve GA and DVP AI performance that is not significantly different from trained users during OS exams.

Methods

- Retrospective analysis compared trained users (TU; n=91) with QC-assisted novice users (NUQC; n=91)
 - Comparability of GA and DVP range distributions assessed with Kolmogorov-Smirnov tests
 - All exams used Lumify C5-2 transducers (Philips Ultrasound LLC, USA)
- QC suite prompted repeat sweeps when probe angle or coupling failed data-quality thresholds
- GA and DVP outputs were assessed against expert biometry and DVP scans to calculate absolute error
- Absolute errors between NUQC and TU groups were compared using Welch’s t-tests

Results

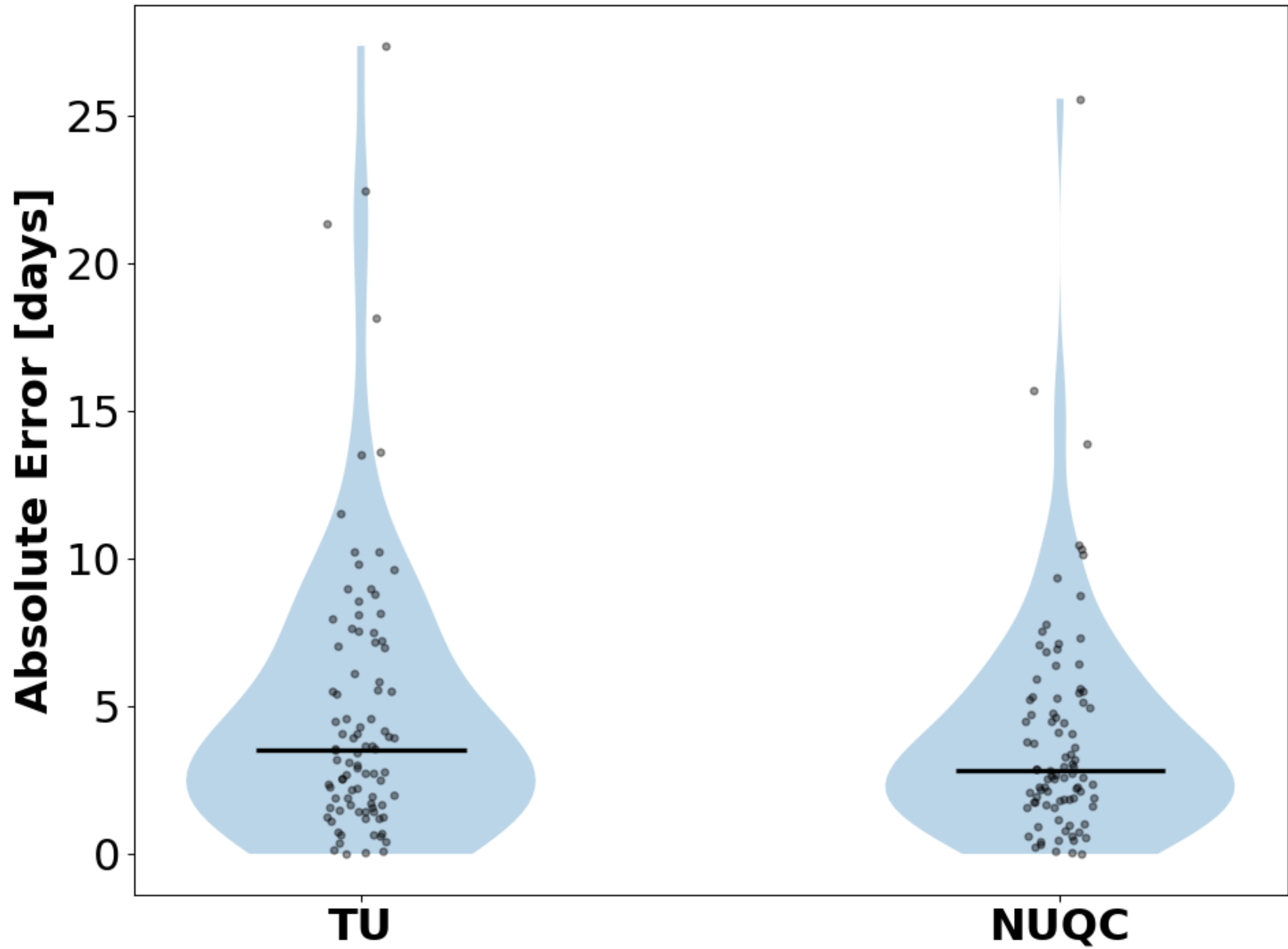


Figure 1. GA absolute error violin plot, TU vs NUQC groups.

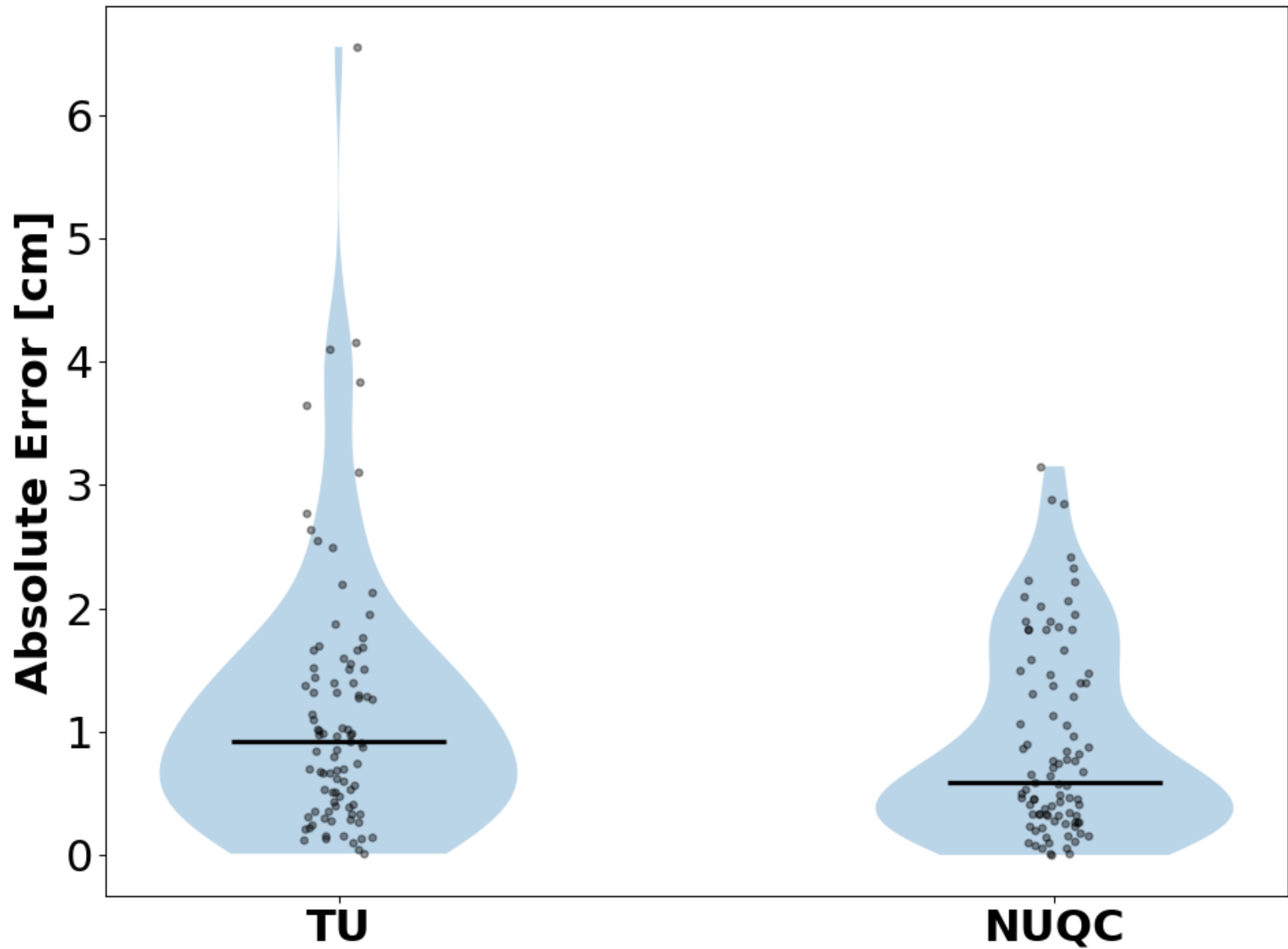


Figure 2. DVP absolute error violin plot, TU vs. NUQC groups.

Results (continued)

	KS Statistic	P-Value
GA	0.12	0.52
DVP	0.11	0.64

Table 1. Kolmogorov-Smirnov distribution test results.

	TU MAE	NUQC MAE	% Error	P-Value
GA	4.93 ± 4.96 days	3.96 ± 3.79 days	-19.7%	0.14
DVP	1.16 ± 1.08 cm	0.89 ± 0.77 cm	-22.6%	0.06

Table 2. Welch’s t-test results between TU and NUQC groups.

Discussion

- GA and DVP distributions were comparable between TU and NUQC cohorts (Table 1)
- QC-assisted novices showed lower MAE for both measures, which was not statistically significant (Figures 1-2, Table 2)
- Novices averaged 0.96 QC errors per exam

Conclusion

- QC-assisted novices performed similarly to trained users
- Performance trended toward lower error in the QC-assisted group
- Future work:
 - Compare novices with and without QC
 - Test the impact of QC on additional AI-assisted obstetric parameters

Acknowledgements

This work was supported, in whole or in part, by the Gates Foundation INV-030576.