PHILIPS

Auto Measure Abdomen

Fast, more consistent measurements to support efficiency and clinical confidence

Claudia Errico, PhD, Senior Clinical Scientist
Xiang Hui Bai, Senior Image Analytics Scientist
Yishuang Meng, Scientist
Elizabeth Brunelle, Clinical Product Manager
Hua Xie, PhD, Principal Scientist
Philips Ultrasound R&D

This white paper describes development and validation of Philips AI-powered Auto Measure Abdomen, including demographic markers for development and data validation, algorithm design and performance testing, and how automation supports gains in efficiency and enhances clinical confidence.

Introduction

Abdominal measurements are an essential part of routine ultrasound exams[1-2], helping clinicians assess organ size, identify abnormalities and support diagnostic decisions. Traditionally sonographers perform these measurements manually, a process that can be time-consuming and workflow-intensive. Each exam may require numerous caliper placements and repeated adjustments, resulting in considerable variation depending on user experience, measurement technique and image interpretation. This manual variability can impact exam consistency, reproducibility and overall efficiency.

Auto Measure Abdomen relies on artificial intelligence (AI) assistance to automatically pre-populate measurement calipers to perform kidney and spleen measurements **(Figure 1)***

Auto Measure Abdomen is designed to automate the following abdominal measurements:

1. **Longitudinal length of the right and left kidney[3-11]**
2. **Transverse width of the right and left kidney[3-11]**
3. **Transverse height of the right and left kidney[3-11]**
4. **Longitudinal length of the spleen[11-15]**

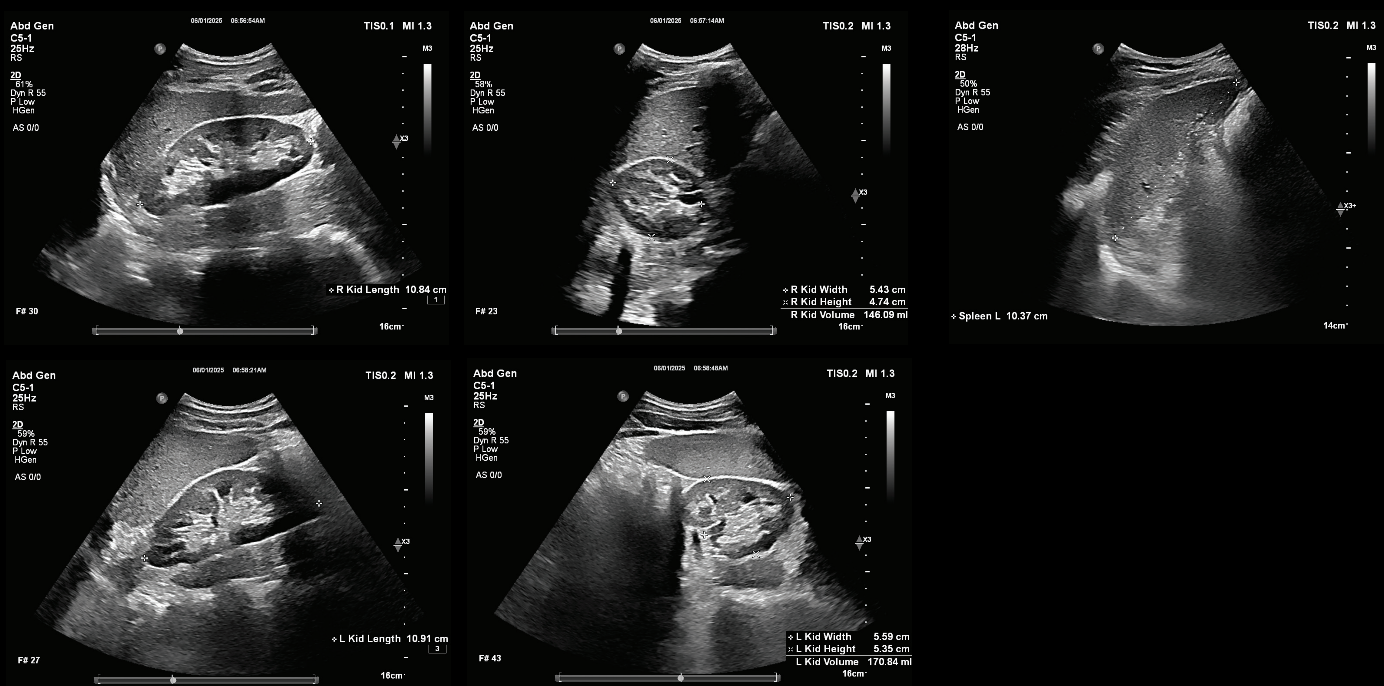


Figure 1
Automated measurements of the length of the longitudinal right and left kidney, width and height of the transverse right and left kidney, and length of the longitudinal view of the spleen.

Data description

Development data were obtained retrospectively from multiple medical institutions across the world.[16] Two sites were located in North America (Northeast and Pacific Northwest), one site in Europe and one site in Asia. The patient population represents various races and is comprised of healthy individuals, as well as patients having indications for receiving abdominal or renal ultrasound.

More than 1,400 subjects in total were used to develop the Auto Measure Abdomen feature (>1,400 subjects for kidney auto measure development, >1,200 subjects for spleen auto measure development) **(Table 1)**.

The AI algorithm was trained on a set of data/patients that is independent from the test and validation data used to evaluate the model performance. The data used for algorithm validation of Auto Measure Abdomen was independent from those used for algorithm development.[16,17] Multiple medical institutions and temporal separation made the data groups independent. Ground truth (GT) was explicitly or implicitly obtained by board-certified sonographers and radiologists. More precisely, the annotation of each development frame originates either from an explicit expert annotation or is derived from a clinical measurement performed on the data as part of standard of care during abdominal and/or renal ultrasound examination.

*Check with your local Philips representative for information about additional measurements that may be available in your geographic region.

Distribution of demographic markers for development and validation data

The data collection includes diverse demographics and varied health statuses to ensure comprehensive inclusion of eligible cases that are representative of the intended use population.[16,17]

Attribute	Development data Kidney 1,454 subjects	Development data Spleen 1,273 subjects	Validation data pool Kidney & spleen 150 subjects
Age years (mean, SD)	50.77 (14.78)	(50.74, 14.75)	58.59 (14.70)
BMI kg/m² (mean, SD)	26.63 (6.04)	(26.56, 5.98)	26.20 (4.74)
Gender **Male% – Female%**	Male 51.03% – Female 48.97%	Male 50.74% – Female 49.26%	Male 50% – Female 50%
Race **White%** **Asian%** **Black%** **Other %**	 18.82% 39.37% 12.39% 29.42%	 20.79% 43.11% 13.94% 22.16%	 64.00% 31.33% 3.3% 0.67%
Pathology **Yes% – No%**	Yes 21.71% – No 78.29%	Yes 15.13% – No 84.87%	**Kidney** Yes 22.00% – No 78.00% **Spleen** Yes 1.33% – No 98.67%

Table 1
Data demographics and distribution for development vs. test and validation data.

Algorithm design and performance testing

The Auto Measure Abdomen algorithm works from input ultrasound image to output measurement calipers. The input image feeds into a convolutional neural network (CNN) which is designed to output a group-predicted caliper position.[17]

The major limitation of ultrasound measurements is user variability because measurements are mostly performed manually. This variability is well known to the clinical community. To estimate our performance target (i.e., the agreement between AI output and manual measurements), we implemented a Bland-Altman analysis,[18,19] which assesses the difference between two observers performing the same measurement. The acceptance criteria for the Auto Measure Abdomen performance were defined using limits of agreement (LoA) from the Bland-Altman analysis, derived per measurement type.

The output of the Auto Measure Abdomen was evaluated for clinical performance in comparison to GT (manual measurement in standard of care and generated by clinical experts).

Table 2 reports the clinical performance of Auto Measure Abdomen AI results with respect to the GT, defined as the average manual measurements done by three clinical experts on validation data. For Auto Measure Abdomen performance, we considered the two measurements (AI vs. GT) to be in agreement if the upper and lower limit of the 95% confidence interval (CI) of the LoA **(Table 2 and Figures 2-5)** were within set boundaries of the maximum allowable difference (±MAD, ±Δ).The MADs were derived by averaging the clinical variability of each measurement obtained via published literature[3-15] and results of an internal study for user variability assessment.

Measurement	Mean difference ± SD	95% CI of LoA	Acceptance criteria (MAD)
Kidney longitudinal length	0.46% ± 3.51%	(-7.10%, 8.02%)	[-14.3%, 14.3%]
Kidney transverse width	0.26% ± 8.79%	(-18.77%, 19.29%)	[-33.7%, 33.7%]
Kidney transverse height	0.54% ± 6.36%	(-13.22%, 14.30%)	[-30.1%, 30.1%]
Spleen longitudinal length	2.34% ± 4.90%	(-8.63%, 13.32%)	[-15.9%, 15.9%]

Table 2
Measurement performance results: Auto Measure Abdomen AI vs. manual results and acceptance criteria.

Because the Bland-Altman LoA between Auto Measure Abdomen and GT meets the performance target, it is demonstrated that Auto Measure Abdomen provides clinically acceptable measurements to the users.

Figure 2
Bland-Altman plot of differences between ground truth and fully automated measurement for kidney sagittal length.

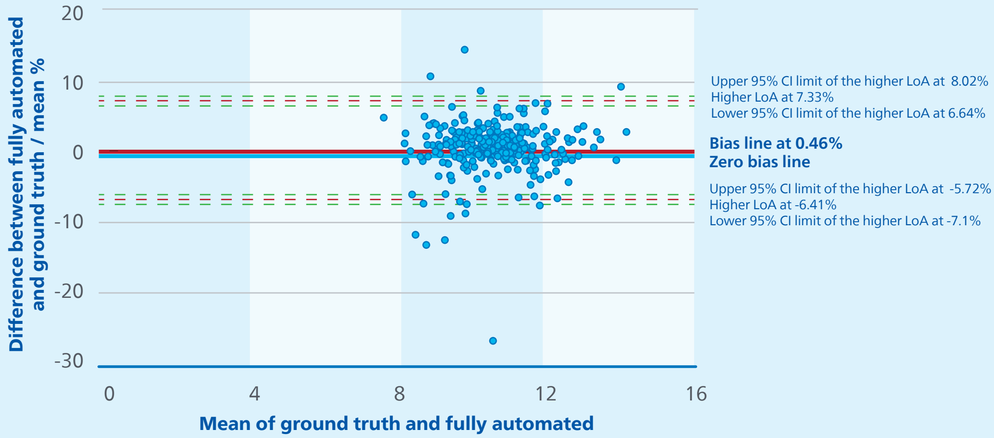


Figure 3
Bland-Altman plot of differences between ground truth and fully automated measurement for kidney transverse width.

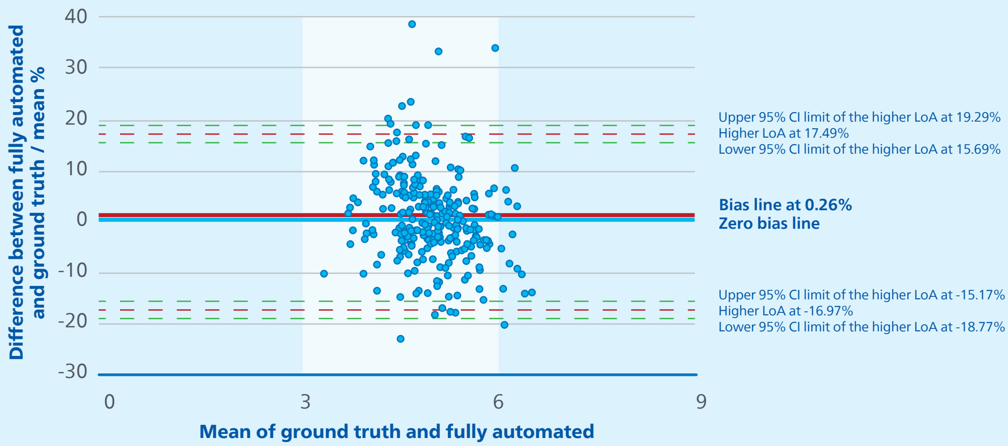


Figure 4
Bland-Altman plot of differences between ground truth and fully automated measurement for kidney transverse height.

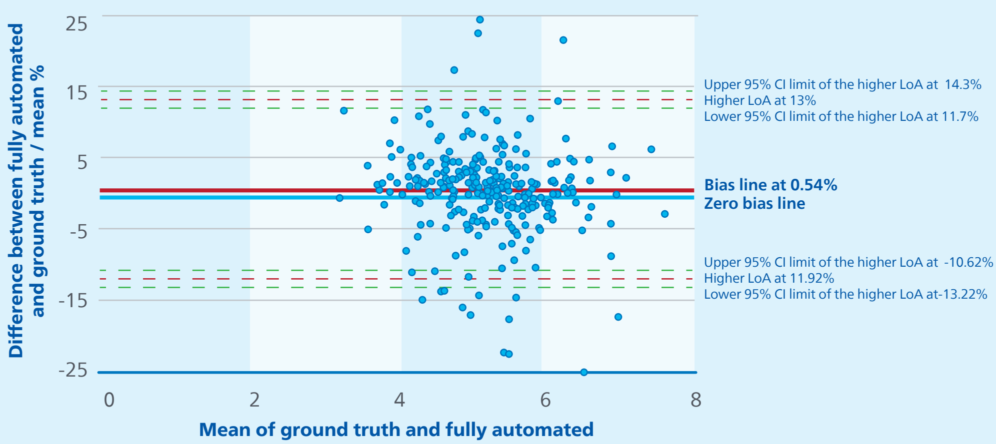
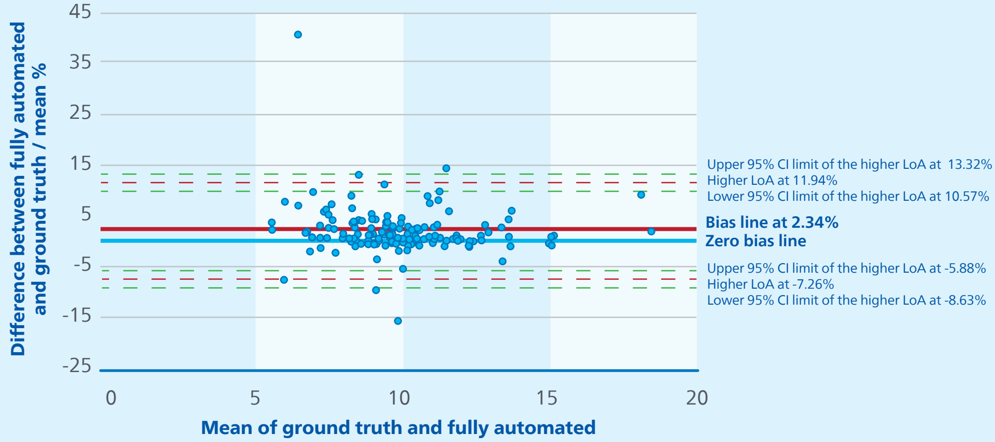


Figure 5
Bland-Altman plot of differences between ground truth and fully automated measurement for spleen length.



Measurement	Accuracy (95% CI)
Kidney longitudinal length (292)	97.68% (97.39%, 97.98%)
Kidney transverse width (271)	93.30% (92.55%, 94.06%)
Kidney transverse height (271)	95.59% (95.03%, 96.14%)
Spleen longitudinal length (145)	96.70% (95.85%, 97.54%)

Table 3
Accuracy for clinical measurements between manual measurements (ground truth) and automated measurements

The results of our study[20,21] demonstrated high agreement of Auto Measure Abdomen AI algorithm-generated measurements with manual measurements by clinical experts (GT) by producing confidence intervals for the limits of agreement (LoA) that met predefined acceptance criteria (**MAD, Table 2**).

The overall Auto Measure Abdomen accuracy (**Table 3**) was demonstrated to achieve an average of over 93% accuracy compared to manual measurements by clinical experts.[20]

Additionally, a quantitative assessment of the caliper placement was calculated to compare the AI model prediction against three clinical experts. **Table 4** shows the agreement between AI and manually edited auto measurements from experts. As demonstrated, the difference is small and the caliper placement (with editing) still meets the acceptance criteria.[21]

Outcomes comparison	Mean difference ± SD(n)	95% CI of mean difference	95% CI of the LoA
Kidney longitudinal length	-0.15% ± 2.26% (292)	-0.41%, 0.11%	-5.04, 4.73%
Kidney transverse width	0.64% ± 3.90% (271)	0.18%, 1.11%	-7.80%, 9.08%
Kidney transverse height	0.32% ± 2.79% (271)	-0.02%, 0.65%	-5.72%, 6.35%
Spleen longitudinal length	0.79% ± 3.02% (145)	0.30%, 1.29%	-5.97%, 7.55%

Table 4
Caliper performance report: Agreement between AI-generated output and manually edited measurement from experts.

The caliper acceptability rate during the review of Auto Measure output is >77% across all measurements. **Table 5** shows the mean difference between Auto Measure Abdomen output and clinical experts vs. mean difference across all experts. The clinical variability seen among experts is comparable to the variability seen between the AI output and experts.[21]

Measurements (N)	AI vs. expert Mean difference (cm)	Among three experts Mean difference (cm)
Kidney sagittal length (1,752)	0.38 (0.36, 0.40)	0.41 (0.39, 0.43)
Kidney transverse width (1,626)	0.87 (0.83, 0.90)	0.86 (0.83, 0.90)
Kidney transverse height (1,626)	0.85 (0.81, 0.89)	0.79 (0.75, 0.83)
Spleen length (870)	0.57 (0.54, 0.61)	0.56 (0.52, 0.60)

Table 5
Mean differences for caliper placement between AI and an expert and across experts.

Turning automation into clinical efficiency

The goal of AI-assisted measurements is to automatically pre-launch the calipers for the user in the anatomical location where typical abdominal measurements are performed (i.e., kidney length, width and height, and spleen length). Auto Measure Abdomen helps users to reduce manual interactions with the system, while - if necessary - speeding up the editing workflow. A caliper will always remain active to support the editing workflow if needed.

The current manual workflow for abdominal measurements is summarized in **Figure 6**.[22]

Manual measurement

Number of keystrokes to perform measurements varies from 3-4, depending on workflow.

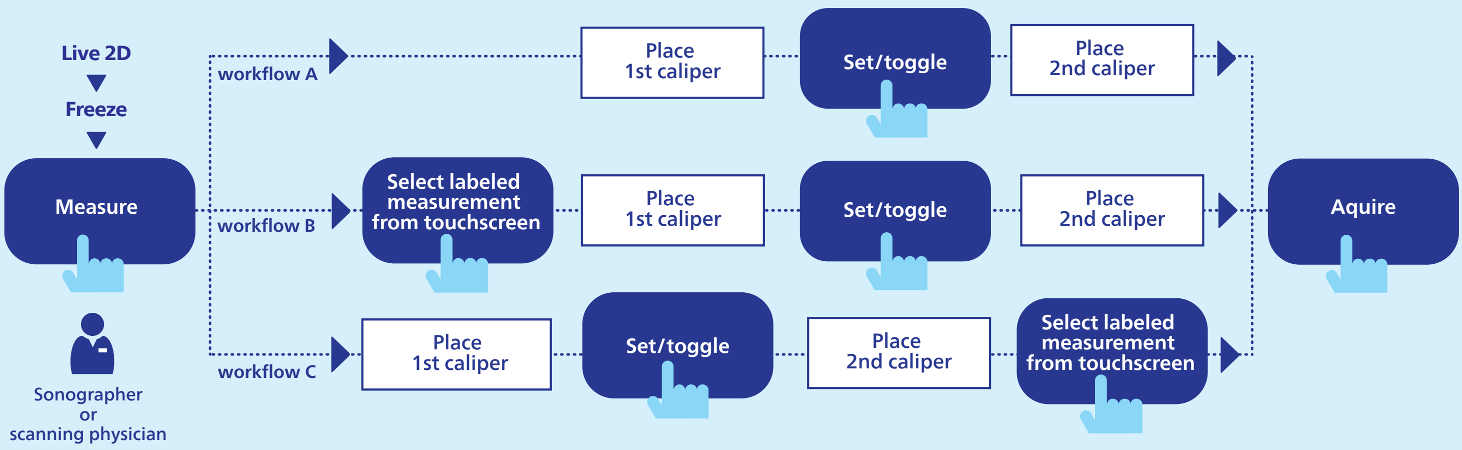


Figure 6
Abdominal measurement manual workflow.

While the workflow can vary from the sonographer vs. scanning physician model or between medical institutions that prefer labeled vs. generic distance measurements, the number of button clicks and manual interaction between the user and the trackball (to move/place the calipers on the desired location within the anatomy) is very repetitive.

Additionally, depending on the site protocol, to complete a full abdominal study with all required abdominal measurements, users might need to repeat these steps up to 12+ times (if we simply consider routine 2D distance measurements required by American College of Radiology [ACR] guidelines).

With Auto Measure Abdomen, an asterisk is added to the labeled measurement where Auto Measure Abdomen is available, and the workflow to perform labeled measurements becomes straightforward with a single button click (**Figure 7**), eliminating the need to place multiple calipers manually and have multiple interactions with the trackpad. The proposed automated workflow for abdominal measurements is summarized in **Figure 7**.

Auto Measure greatly simplifies workflow

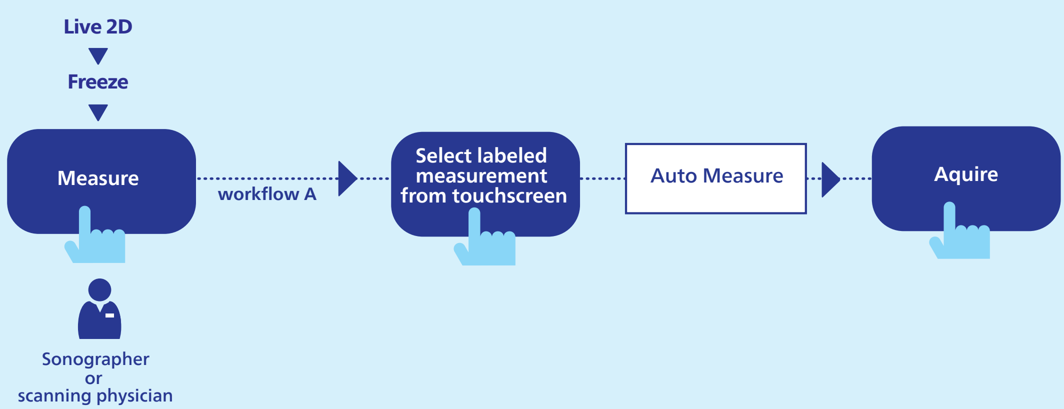


Figure 7
Abdominal measurement automated workflow with AI Auto Measure Abdomen.

To demonstrate reduction in keystrokes, the button pushes between manual measurement and fully automated measurement with the Auto Measure Abdomen feature were compared. When comparing the two methods (i.e., button clicks highlighted in **Figure 6 and Figure 7**, if a single measurement is performed on a frame (e.g., kidney length and spleen length), automation can lead to a 25% decrease in keystrokes. Whereas, in a scenario where multiple measurements are done on the same image (i.e., kidney width and height), then the number of button pushes is reduced up to 33%.[22]

The integration of advanced AI capabilities into ultrasound imaging marks a transformative shift in how clinicians acquire and interpret data. In support of Auto Measure Abdomen performance, we gathered clinical perspective from two experienced sonographers from Boston Medical Center. We are pleased to feature insights from Gretchen Sammy, ARDMS-registered sonographer, Ultrasound Manager with 20 years of experience, and Erin Walsh Irwin, ARDMS-registered sonographer, Ultrasound Supervisor with 15 years of experience, whose clinical expertise and practical perspectives enrich the findings presented in this paper. They commented on how they use Auto Measure Abdomen to boost efficiency, enhance accuracy and protect their team, saying:

"The Philips Auto Measure tool has become a game-changer for our ultrasound workflow. By automating key measurement tasks, it allows our sonographers to reduce scanning time by up to 30%, without sacrificing clinical precision. In our evaluations, Auto Measure consistently placed measurements exactly where we would expect them – saving time while maintaining the accuracy we rely on."

"When adjustments are needed, the tool's intuitive design makes fine-tuning quick and effortless. This ease of use directly translates to fewer keystrokes, helping reduce repetitive motions that contribute to workplace injuries – an essential consideration for any sonography department."

> " Philips Auto Measure tool has become a game-changer for our ultrasound workflow."
>
> **— Gretchen Sammy and Erin Walsh Irwin**
> **ARDMS-registered sonographers**
> **Boston Medical Center**

"For facilities focused on efficiency, accuracy, and sonographer well-being, I highly recommend integrating tools like Philips Auto Measure Abdomen. It's a smart investment in both your team and your workflow."

In line with our clinical experts, we evaluated time savings, a study was conducted in a clinical setting with multiple sonographers performing manual and automated measurements with Auto Measure Abdomen. Seven sonographers were recruited to perform two exams with manual vs. automated measurements on five subjects. The scanning protocol required each user to acquire seven frames (i.e., abdominal anatomies that currently have auto measurements available) per subject, per exam, for a total of 70 frames.

A total of seven internal expert readers (each certified by the American Registry for Diagnostic Medical Sonography [ARDMS]) independently performed manual measurements, fully automated measurements (accepting AI output without edits) and semi-automated measurements (editing AI output if needed) with the assistance of Auto Measure Abdomen.

The measurement order was randomized for readers between manual measurement and automated measurement with at least 10 minutes of wash-out time between manual and automated measurement to minimize bias.

The time it took to complete the measurement process was logged for each method and showed an average time savings of 55% for fully automated measurements, and an average time savings of 36% with user edits when applied to auto measurements **(Figure 8)**.[22]

Measurement time comparison

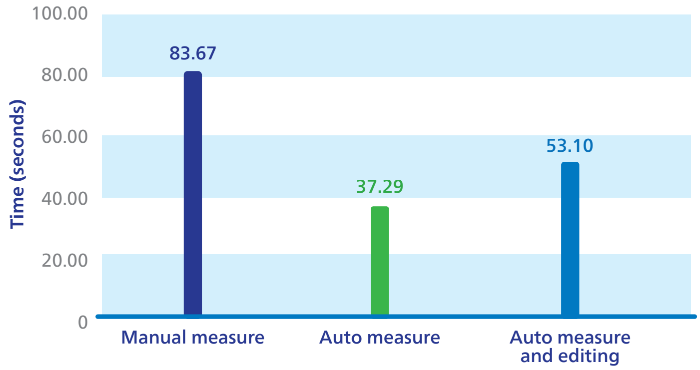


Figure 8
Measurement time comparison showing time reduction by Auto Measure Abdomen with respect to manual measurements.

Conclusion

Auto Measure Abdomen shows strong alignment with expert manual measurements and meets predefined accuracy standards. The feature reduces user variability and streamlines workflow by automatically placing calipers for key abdominal measurements. Clinical evaluations demonstrated significant reductions in keystrokes and over 50% time savings[23] with fully automated use. Supported by diverse development and validation datasets, the algorithm performs reliably across different patient groups. Overall, Auto Measure Abdomen enhances measurement consistency and improves efficiency in abdominal ultrasound exams, while performing within clinical variability established from experts.

References

1. The AIUM Practice Parameter for the Performance of an Ultrasound Examination of the Abdomen and/or Retroperitoneum. J Ultrasound Med. 2022;41(4):E1-E8. doi:10.1002/jum.15874
2. The ACR Medical Image Quality Assessment System (MIQAS): A Unified Approach to Image Quality Assessment in Radiology, 2025 American College of RadiologyNCSS.com., PASS Sample Size Software. Bland-Altman Method for Assessing Agreement in Method Comparison Studies. Chapter 427.
3. Piras D, Masala M, Delitala A, Urru SAM, Curreli N, Balaci L, Ferreli LP, Loi F, Atzeni A, Cabiddu G, Racugno W, Ventura L, Zoledziewska M, Steri M, Fiorillo E, Pilia MG, Schlessinger D, Cucca F, Rule AD, Pani A. Kidney size in relation to ageing, gender, renal function, birthweight and chronic kidney disease risk factors in a general population. Nephrol Dial Transplant. 2020 Apr 1;35(4):640-647.
4. Emamian SA, Nielsen MB, Pedersen JF, Ytte L. Kidney dimensions at sonography: correlation with age, sex, and habitus in 665 adult volunteers. (1993) AJR. American journal of roentgenology. 160 (1): 83-6.
5. Ablett M J, Coulthard A, Lee R E J, et al. How reliable are ultrasound measurements of renal length in adults? [J]. The British journal of radiology, 1995, 68(814): 1087-1089.
6. Geelhoed J J M, Kleyburg-Linkers V E, Snijders S P E, et al. Reliability of renal ultrasound measurements in children[J]. Pediatric nephrology, 2009, 24: 1345-1353.
7. Jeannette Bakker, Marco Olree, Robert Kaatee, et al. Renal Volume Measurements: Accuracy and Repeatability of US Compared with That of MR Imaging. Radiology 1999; 211:623–628
8. Eze C U, Marchie T T, Ohagwu C C, et al. Observer variability in sonographic measurement of kidney sizes among children in Benin City, Nigeria[J]. West Indian med. j, 2013: 817-824.
9. Emamian S A, Nielsen M B, Pedersen J F. Intraobserver and interobserver variations in sonographic measurements of kidney size in adult volunteers: a comparison of linear measurements and volumetric estimates[J]. Acta radiologica, 1995, 36(4): 399-401.
10. Kim DW, Ahn HG, Kim J, Yoon CS, Kim JH, Yang S. Advanced Kidney Volume Measurement Method Using Ultrasonography with Artificial Intelligence-Based Hybrid Learning in Children. Sensors (Basel). 2021 Oct 14;21(20):6846.
11. Singla R, Ringstrom C, Hu R, Hu Z, Lessoway V, Reid J, Nguan C, Rohling R. Automatic measurement of kidney dimensions in two-dimensional ultrasonography is comparable to expert sonographers. J Med Imaging (Bellingham). 2023 May;10(3):034003.
12. Li PS, Ying M, Chan KH, Chan PW, Chu KL. The reproducibility and short-term and long-term repeatability of sonographic measurement of splenic length. Ultrasound Med Biol. 2004 Jul;30(7):861-6.
13. Lucius C, Meier J, Gschmack A, Zervides C, Jenssen C, Dong Y, Graumann O, Petry M, Dietrich CF. Ultrasound of the spleen - an update on measurements, reference values, and influencing factors. A systematic review. Med Ultrason. 2025 Jun 16;27(2):185-194. doi: 10.11152/mu-4436. Epub 2024 Sep 22. PMID: 39420823.
14. Chow K U, Luxembourg B, Seifried E, et al. Spleen size is significantly influenced by body height and sex: establishment of normal values for spleen size at US with a cohort of 1200 healthy individuals[J]. Radiology, 2016, 279(1): 306-313.
15. Hosey RG, Mattacola CG, Kriss V, Armsey T, Quarles JD, Jagger J. Ultrasound assessment of spleen size in collegiate athletes. Br J Sports Med. 2006 Mar;40(3):251-4; discussion 251-4.
16. Internal documentation: Validation report D002210029.
17. Internal documentation: Validation report D002209622.
18. Lu MJ, Zhong WH, Liu YX, Miao HZ, Li YC, Ji MH. Sample Size for Assessing Agreement between Two Methods of Measurement by Bland-Altman Method. Int J Biostat. 2016 Nov 1;12(2)
19. Francq BG, Berger M, Boachie C. To tolerate or to agree: A tutorial on tolerance intervals in method comparison studies with BivRegBLS R Package. Stat Med. 2020 Dec 10;39(28):4334-4349.
20. Internal documentation: Validation report D002351044.
21. Internal documentation: Validation report D002383552.
22. Internal documentation: Validation report D002351150.
23. Internal documentation. Validation report D001785669.

CE
2797

PHILIPS

www.philips.com

Printed in the Netherlands.
00001583-00-01 * JUL 2026