

Geometric, acquisition and image-formation limits of two-dimensional ultrasound volumetry across simulated fluid-filled structures

Wesley da Silva Oliveira

wesleyoliveira@usp.br

Instituto de Pesquisas Energéticas e Nucleares

Research Article

Keywords: Ultrasonography, Volumetry, Virtual imaging trial, Measurement uncertainty, Quantitative ultrasound, Acquisition geometry

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10911309/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Geometric, acquisition and image-formation limits of two-dimensional ultrasound volumetry across simulated fluid-filled structures

Wesley da Silva Oliveira

Instituto de Pesquisas Energéticas e Nucleares (IPEN-CNEN), São Paulo, Brazil

Corresponding author: Wesley da Silva Oliveira; wesleyoliveira@usp.br

ORCID: <https://orcid.org/0009-0002-2078-1193>

Abstract

Two-dimensional ultrasound volume estimates are often derived from a few linear dimensions and a fixed geometric coefficient, so their accuracy depends on shape assumptions and acquisition. We used a virtual imaging trial to separate intrinsic geometric error from acquisition-plane and image-formation effects and to test whether formula performance changed across fluid-filled cavity morphologies. The primary bladder-like design included 2,048 exact geometries (G), 24,576 acquisition-perturbed measurements (A), and 256 prospectively stratified synthetic-ultrasound acquisitions (U), of which 251 were measurable. Three conventional formulas were compared using absolute percentage error (APE), with >15% prespecified as the primary failure threshold. In the bladder-like design, median APE for the 0.52 formula increased from 12.57% in G to 14.49% in A and 16.29% in U; operational U failure was 54.3%. The 0.2π and 0.625 formulas had lower U median APEs (8.23% and 7.75%), but their relative performance changed across simple-cyst-like, gallbladder-like, and dilated renal collecting-system-like extensions. Image formation produced submillimetric mean dimensional shifts, whereas single-frame field-of-view limitation created a separate 2.39% weighted unimageability mechanism. Fixed-coefficient 2D ultrasound volumetry therefore carries geometry-dependent error before image formation, with additional acquisition and imaging effects. No evaluated formula was uniformly optimal across the controlled morphologies.

Keywords Ultrasonography; Volumetry; Virtual imaging trial; Measurement uncertainty; Quantitative ultrasound; Acquisition geometry

Introduction

Two-dimensional (2D) ultrasound volumetry is widely used because it is fast, accessible, and easy to repeat. In routine practice, volume is commonly estimated from a few orthogonal linear dimensions multiplied by a fixed geometric coefficient. The calculation is convenient, but it also embeds a shape assumption: identical measured dimensions can represent different true volumes when morphology changes. The formulas evaluated here use coefficients of 0.52, 0.2π , and 0.625, with the 0.625 formulation averaging longitudinal and transverse depth.

The final estimate can be affected at several stages. A formula may be geometrically inaccurate even when the true maximal dimensions are known; the acquisition plane may miss the maximal section; and image formation may shift the apparent fluid-tissue interface. Clinical and phantom measurements usually contain these effects simultaneously, making their individual contributions difficult to identify. The distinction matters because changing a coefficient can correct average bias, whereas errors related to morphology, plane selection, field of view, or interface localization require different solutions.

High measurement repeatability does not necessarily imply agreement with a reference volume. In three-dimensional ultrasonographic muscle-volume assessment, Ritsche et al. reported high reliability alongside systematic differences from MRI that varied with muscle shape and location [1]. Virtual imaging trials provide a complementary way to examine such discrepancies because geometry, acquisition conditions, imaging physics, and ground truth can be controlled independently [2].

Accordingly, we designed a nested virtual imaging trial with three levels: exact geometry (G), acquisition-perturbed geometry (A), and synthetic ultrasound (U). We hypothesized that coefficient choice alone would not determine volumetric validity and that morphology and plane selection would materially alter error. A bladder-like family formed the primary experiment, followed by a prospectively defined transferability extension with simple-

cyst-like, gallbladder-like, and dilated renal collecting-system-like cavities. The study was designed to identify when and why 2D volumetry exceeds prespecified error limits under controlled morphology and acquisition conditions; it was not designed to estimate organ-specific clinical failure prevalence.

Methods

Study design

This fully computational virtual imaging trial used only synthetic geometries, predefined acquisition planes, and simulated B-mode data. Study design, endpoints, formula set, acquisition perturbations, sampling rules, and quality-control thresholds were specified before definitive outcomes were generated. The virtual populations were controlled design spaces rather than epidemiologic samples. Cases were not removed or replaced because of unfavorable volumetric performance, and the primary APE threshold was not changed after results were available.

The three study layers were sequential. G represented formula error when exact cavity geometry and centered orthogonal planes were known. A applied controlled plane rotation and offset to the same geometry without image formation. U retained the same acquisition definitions and added synthetic B-mode formation plus physics-limited local interface localization. We interpreted the layers as successive stages of measurement rather than independent error components to be added together.

Primary bladder-like geometry population

The definitive bladder-like population contained 2,048 geometries generated with a scrambled Sobol low-discrepancy design. Target internal-fluid volume was log-uniform from 10 to 500 mL. Morphology varied independently across length-to-width ratio (1.10-2.40), width-to-height ratio (1.00-1.60), longitudinal and transverse superelliptic shape exponents (1.70-3.00), pearness (-0.15 to 0.35), caudal neck taper (0-0.30), left-right asymmetry (-0.12 to 0.12), and dorsoventral asymmetry (-0.10 to 0.10). The geometry used superelliptic transverse sections whose centerlines could be translated asymmetrically without changing local cross-sectional area. Each object was isotropically rescaled so that the analytically integrated internal volume matched its assigned target volume. Pathologies such as stones, sediment, gas, masses, and inflammatory wall thickening were intentionally excluded.

For G, the longitudinal plane was centered at $y=0$ and the transverse plane at $x=0$. Exact cavity-plane intersections yielded length (L), longitudinal depth (D_L), width (W), and transverse depth (D_T). For formulas using a single depth, D_T was the frozen depth definition.

Acquisition perturbation layer

Each bladder-like geometry contributed 12 A acquisitions: one ideal acquisition and 11 perturbed acquisitions, for 24,576 measurements. Longitudinal and transverse planes could rotate independently by -20° to $+20^\circ$ around the dorsoventral axis. Plane offsets ranged from -0.20 to $+0.20$ of the relevant cavity radius. Compound angle-plus-offset errors were allowed. Exact plane-body intersections were computed numerically with 4,097 horizontal samples and boundary interpolation; the same frozen plane definitions were later used by U.

Volumetric formulas

Three conventional formulas were evaluated. The 0.52 and 0.2π formulas use the same $L \times W \times D_T$ dimensional product and differ only in the fixed coefficient. The 0.625 formula uses the average of longitudinal and transverse depth.

$$V = 0.52 \times L \times W \times D_T$$

$$V = 0.2\pi \times L \times W \times D_T$$

$$V = 0.625 \times L \times W \times \frac{D_L + D_T}{2}$$

Synthetic-ultrasound layer

The definitive U layer used PyMUST version 0.1.9 [3], a Python implementation of the MUST/SIMUS framework. The underlying SIMUS method has been described and benchmarked against other ultrasound

simulators [4, 5]. Before definitive production, technical cross-simulator gates used k-Wave-based references to confirm submillimetric axial localization in simplified targets and acceptable curved-interface dimensional agreement; these validation runs were not part of the scientific outcome dataset [6].

The final U pipeline used the C5-2v convex probe template at its native center frequency (3.57 MHz), 128 elements, 0.508-mm pitch, 49.57-mm radius, and 79% bandwidth. A single zero-delay transmission was beamformed on a 256×256 polar grid using convex delay-and-sum processing with automatic receive f-number. B-mode dynamic range was 30 dB. Diffuse tissue scatterers were simulated at 5.882 scatterers/mm² with random reflection coefficients; scatterers were removed from the internal fluid cavity. No explicit echogenic wall was added, so the measurement target was the fluid-tissue interface rather than a modeled wall thickness.

Wall positions were obtained with physics-limited, target-aware local localization rather than an autonomous detector. For each expected wall, the algorithm evaluated the positive outward B-mode gradient within a ±4-mm truth-centered window, after 0.7-mm Gaussian smoothing, and used the gradient centroid as the localized interface. This design isolates image-formation and local interface-position error; it does not model complete clinical observer behavior.

Because high-fidelity U generation was computationally intensive, the U sample was prospectively reduced before definitive outcomes to 256 acquisitions from the already frozen pre-outcome reference subset. Sampling was exactly stratified across four volume quartiles and four acquisition strata (ideal plus three severity levels), with 16 acquisitions per cell. Overall U estimates therefore used frozen design weights. Five acquisitions were prospectively classified as single-frame field-of-view (FOV) unimageable before SIMUS execution and were retained without replacement. Conditional U accuracy used measurable acquisitions only; operational failure counted either FOV unimageability or measured APE above the relevant threshold.

Technical QC required no image-edge hits, wall-localization absolute error ≤2.0 mm, median dimension error ≤1.5 mm, maximum dimension error ≤3.0 mm, center error ≤2.0 mm, and fan-support distance ≤2.0 mm. Three acquisitions slightly exceeded the prespecified 2.0-mm wall-localization criterion (2.034-2.238 mm). Their original measurements were retained without rerunning, replacement, manual correction, or threshold relaxation, and sensitivity analyses evaluated their influence.

Prospective transferability extension

Three anatomy-inspired fluid-cavity families were prospectively added after the primary design to test whether morphology-dependent behavior persisted beyond the bladder-like geometry. These were engineering design spaces, not clinical cases or disease-prevalence models. The simple-cyst-like family was an exact triaxial ellipsoid (1-300 mL) motivated by the smooth, thin-walled morphology of simple renal cysts and the common use of ellipsoid volume approximations [7-9]. The gallbladder-like family (5-120 mL) introduced elongation, pearness, neck taper, and mild asymmetry in a deliberately broad anatomy-inspired design space informed by sonographic gallbladder literature [10, 11]. The dilated renal collecting-system-like family (1-65 mL) included pelvis-dominant and communicating pelvicalyceal modes, with funnel taper and optional smooth lobulation informed by ultrasound volumetry of hydronephrosis [12, 13].

Each extension family contained 512 G geometries and 12 A acquisitions per geometry (6,144 A measurements). The original prospective U plan selected 64 acquisitions per family across the same 16 volume-by-acquisition cells. Because the extension was designed as a transferability check rather than three additional full quantitative experiments, a prospective sampling amendment was made before any definitive extension U outcome existed. The amendment retained the pre-existing within-cell index 0 case, leaving one case per cell and 16 U acquisitions per family (48 total). Selection did not use G, A, U, or transportability outcomes. The final extension used the same C5-2v image-formation and localization physics and class-specific frozen design weights.

Outcome measures and statistical analysis

The primary metric was absolute percentage error, defined as $APE = 100 \times \frac{|V_{est} - V_{true}|}{V_{true}}$. Primary failure was prespecified as APE >15%, with APE >10% as a secondary threshold. These thresholds were used as comparative measurement-error tolerances for the controlled virtual trial and were not intended as universal clinical decision

cutoffs. Signed percentage error (PE), mean absolute error (MAE), root-mean-square error (RMSE), and the effective coefficient $k_{true} = \frac{V_{true}}{L \times W \times D_T}$ were secondary measures.

G and A were summarized over their complete controlled designs. U estimates were design-weighted because of prospective stratification. For measurable U acquisitions, conditional failure was the weighted fraction above threshold. Operational failure used the full design-weight denominator and counted prespecified FOV-unimageable cases as failures. Formula performance was not interpreted as an epidemiologic rate.

Agreement was evaluated on the log-ratio scale $\ln(\frac{V_{est}}{V_{true}})$, with geometric bias and limits of agreement obtained by exponentiating the mean log ratio and mean ± 1.96 SD, following the general agreement framework of Bland and Altman [14]. Heteroscedasticity was described by Spearman associations of true volume with absolute error and with APE, and by the log-log slope of absolute error versus true volume. Morphology correlations were descriptive because the virtual populations were low-discrepancy controlled designs rather than probability samples.

The frozen analysis plan named first-order and total-order Sobol indices. Direct variance-based Sobol indices were not estimated because the archived definitive design consisted of a single scrambled-Sobol space-filling sample rather than the paired A/B and pick-freeze matrices required by standard first-order and total-order estimators. A post hoc surrogate was not added; factor relationships were therefore reported descriptively from the locked design. A required sensitivity analysis addressed the three preserved U QC exceptions by (i) excluding them with weight renormalization and (ii) conservatively forcing them to count as failures for every formula.

AI tools were used solely to support translation and to check bibliographic references. The author reviewed and verified all text, references, and scientific content.

Additional design ranges, agreement diagnostics, QC sensitivity analyses, and supplementary figures are provided in Online Resource 1.

Table 1 Frozen study architecture and sample sizes

Component	Sample	Controlled variation	Scientific role
Primary bladder-like G	2,048 geometries	Centered exact orthogonal planes	Intrinsic geometry/formula error
Primary bladder-like A	24,576 acquisitions	1 ideal + 11 angle/offset perturbations per geometry	Adds acquisition-plane error
Primary bladder-like U	256 acquisitions (251 measured)	Prospectively stratified synthetic B-mode; 5 FOV-unimageable retained	Adds image formation/local interface localization and FOV
Simple cyst-like extension	512 G; 6,144 A; 16 U	Near-ellipsoidal reference family	Transferability
Gallbladder-like extension	512 G; 6,144 A; 16 U	Elongation, pearness, taper, asymmetry	Transferability
Renal collecting-system-like extension	512 G; 6,144 A; 16 U	Pelvis-dominant or communicating pelvicalyceal morphology	Transferability

Results

Study completion

The primary bladder-like virtual imaging trial contained 2,048 G geometries and 24,576 A measurements. The prospectively stratified U layer contained 256 acquisitions; 251 were measurable and five remained single-frame FOV unimageable without replacement. The three retained wall-localization QC exceptions were acquisition IDs 2780, 6006, and 22644. The transferability extension contained 512 G geometries and 6,144 A acquisitions per geometry family and 16 U acquisitions per family (48 total); all 48 were measurable.

Primary bladder-like experiment

Substantial volumetric error preceded image formation (Table 2; Figs. 1 and 2). At G, median APE was 12.57% with the 0.52 formula, 7.74% with 0.2π , and 7.75% with the 0.625 formulation. Failure above 15% occurred in 38.9%, 24.9%, and 26.5% of geometries, respectively. Mean signed PE was -11.12% for 0.52, +7.39% for 0.2π , and +8.59% for 0.625.

At A, median APE changed to 14.49%, 7.50%, and 6.97%, with >15% failure in 47.5%, 19.7%, and 17.6%. Among the 22,528 non-ideal acquisitions, perturbation increased APE relative to the corresponding G measurement in 60.7%, 45.6%, and 41.8% for the three formulas. Acquisition therefore acted in both directions: it worsened some pre-existing geometric errors while partially compensating others.

In design-weighted U among measurable acquisitions, median APE was 16.29% for 0.52, 8.23% for 0.2π , and 7.75% for 0.625. Conditional >15% failure was 53.2%, 22.8%, and 18.7%. Including FOV unimageability in the denominator increased operational failure to 54.3%, 24.7%, and 20.6%.

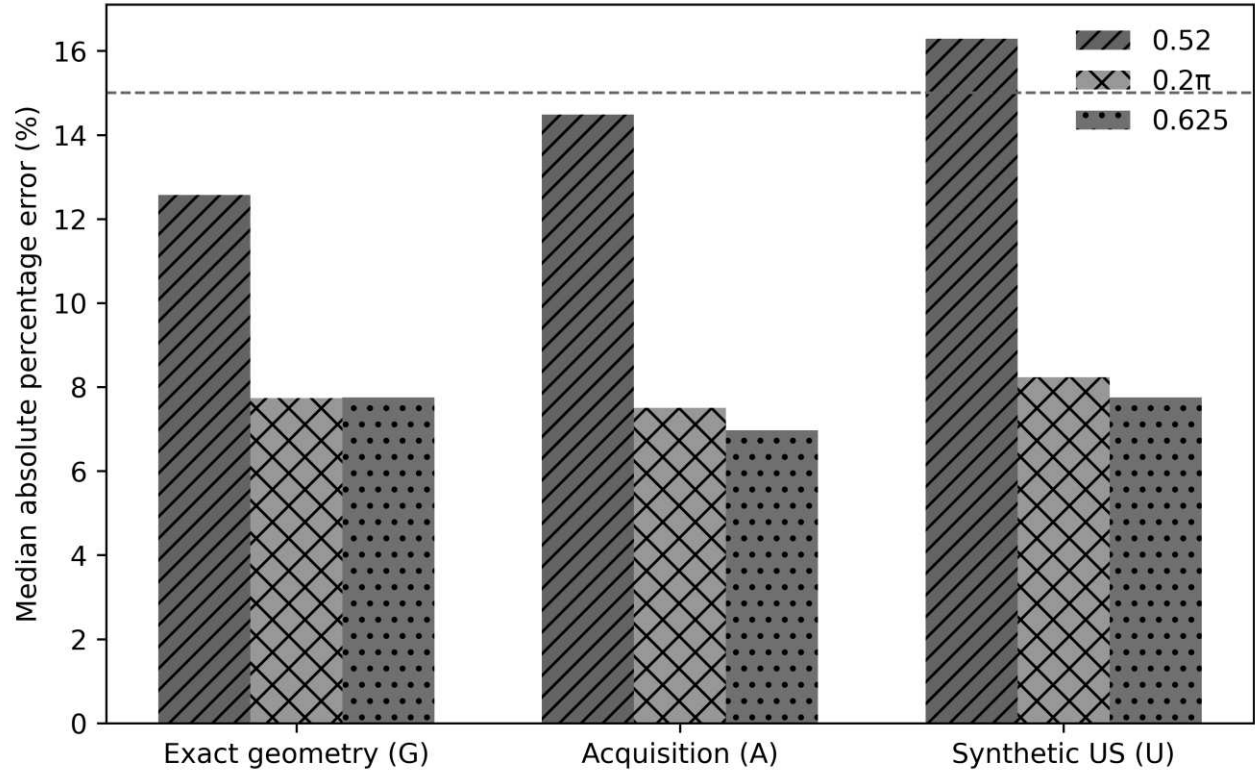


Fig. 1 Absolute percentage error across the nested primary bladder-like experiment. Bars show median APE for each fixed 2D volumetric formula in exact geometry (G), acquisition-perturbed geometry (A), and design-weighted synthetic ultrasound (U). The dashed horizontal line marks the prespecified 15% primary failure threshold

The design-weighted FOV-unimageability fraction was 2.39%. All five unimageable acquisitions occurred in the largest true-volume quartile (approximately >188 mL), corresponding to 9.55% of that quartile and 0% in the lower three quartiles. FOV limitation therefore constituted a distinct failure mechanism independent of formula error among measurable images.

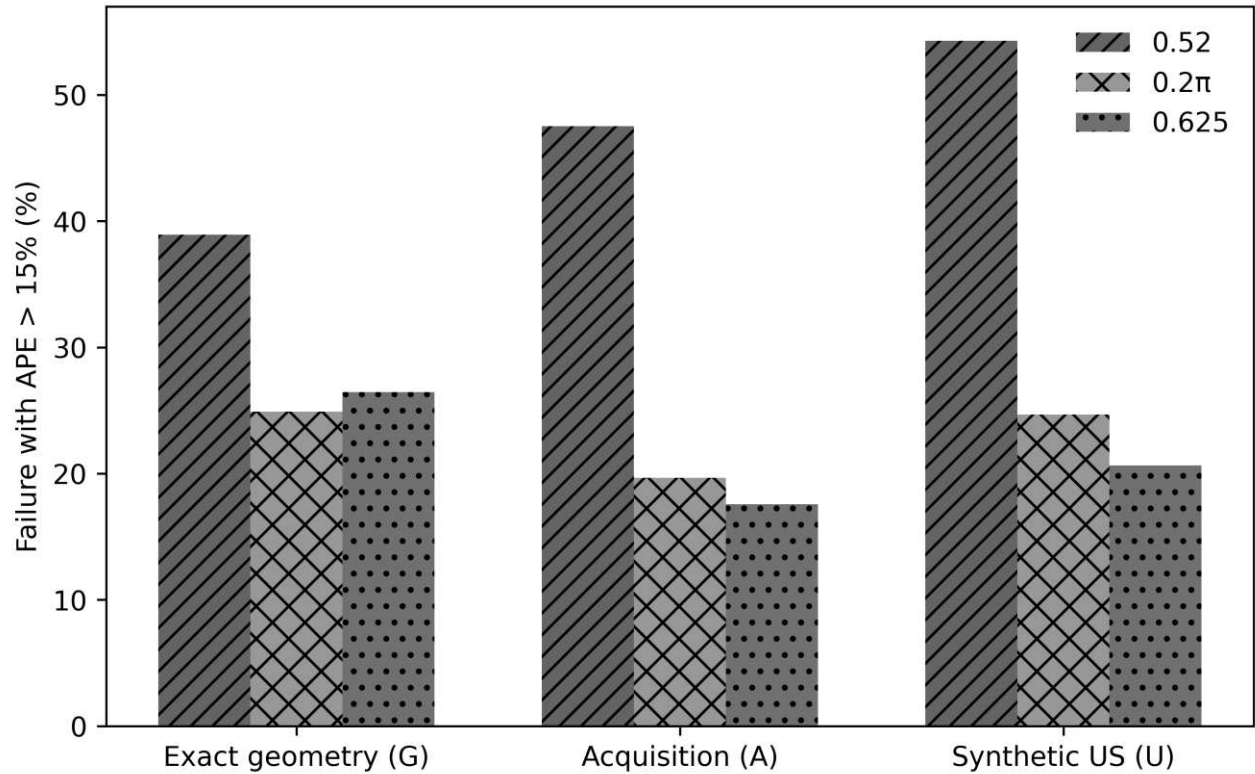


Fig. 2 Failure frequency across the nested primary bladder-like experiment. For U, failure is operational: a case is counted as failed when it is single-frame FOV unimageable or, if measurable, has APE >15%. G and A contain no imageability component

Image formation introduced comparatively small dimension shifts. Weighted mean U-A differences were -0.59 mm for L, -0.33 mm for W, -0.05 mm for D_L , and -0.02 mm for D_T . The corresponding weighted mean changes in APE were +1.31, +0.11, and -0.20 percentage points for 0.52, 0.2π , and 0.625. The image layer therefore added real but smaller error than the geometric and acquisition layers, and its effect on APE depended on the direction of pre-existing bias.

The three preserved QC exceptions did not determine the principal U findings. Excluding all three changed median APE by -0.19, +0.23, and +0.10 percentage points for 0.52, 0.2π , and 0.625, and changed operational failure by -0.50, +0.27, and +0.23 percentage points. Conservatively forcing all three acquisitions to count as failures for every formula yielded operational failure of 54.3%, 25.8%, and 21.7%, respectively.

Effective coefficient, morphology, and multiplicative error

The effective coefficient k_{true} distribution shifted and broadened across $G \rightarrow A \rightarrow U$ (Fig. 3). Median k_{true} was 0.593, 0.608, and 0.620, with 5th-95th percentile ranges of 0.493-0.684, 0.510-0.724, and 0.511-0.760. In G, k_{true} was most strongly associated with longitudinal shape exponent (Spearman $\rho=0.825$; Fig. 6) and transverse shape exponent ($\rho=0.502$); true volume itself had essentially no association ($\rho \approx 0$).

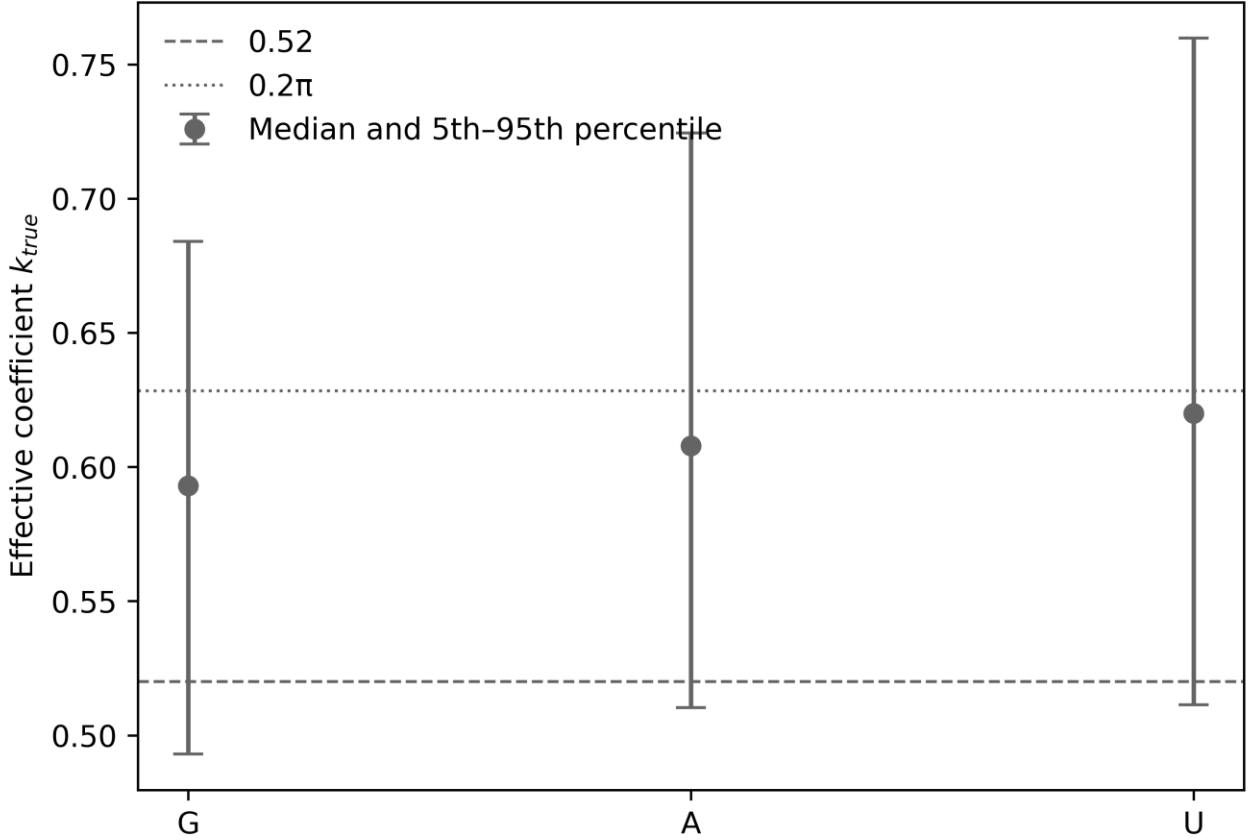


Fig. 3 Distribution of the effective coefficient $k_{true} = \frac{V_{true}}{L \times W \times D_T}$ in the primary bladder-like experiment. Points denote the median and vertical ranges the 5th-95th percentiles. U values are design-weighted among imageable acquisitions. Horizontal reference lines show the fixed 0.52 and 0.2π coefficients applied to the same $L \times W \times D_T$ dimensional product

The 0.52 and 0.2π formulas apply different constants to the identical $L \times W \times D_T$ product. Their log-ratio error distributions therefore had identical dispersion within each stage: log SD 0.101 at G, 0.107 at A, and 0.124 at U. Changing the fixed coefficient shifted bias but could not remove geometry-dependent spread in k_{true} .

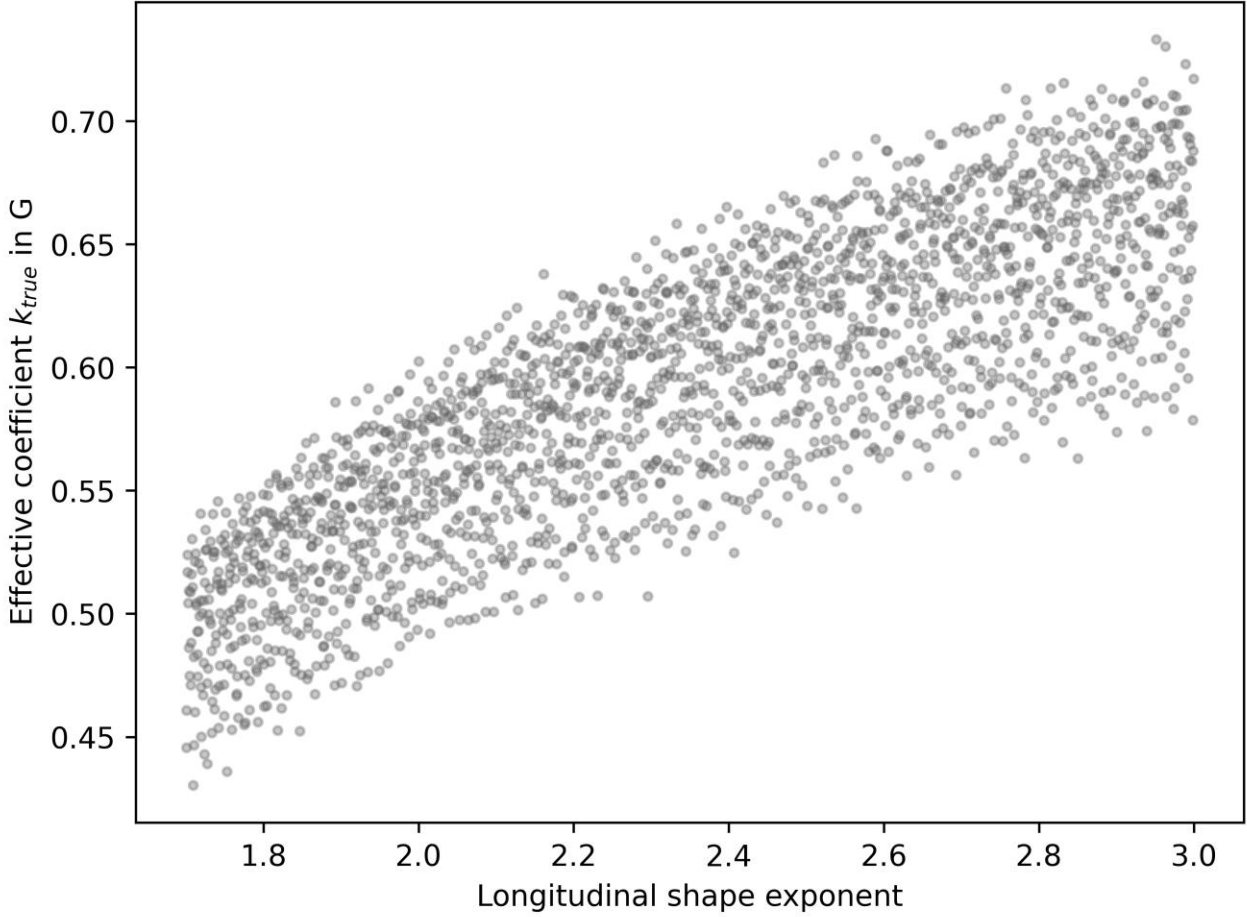


Fig. 6 Relationship between longitudinal shape exponent and exact-geometry effective coefficient in the primary bladder-like population. Correlation is descriptive because the virtual population is a controlled design space rather than an epidemiologic sample

Absolute error in milliliters was heteroscedastic with volume, whereas percentage error was approximately scale-stable. For the 0.52 formula at A, Spearman ρ between true volume and absolute error was 0.825, whereas ρ between true volume and APE was 0.001. The log-log slope of absolute error versus true volume was 1.000 (Online Resource 1, Fig. S1), supporting a predominantly multiplicative error structure.

Prospective transferability extension

The same formulas behaved differently across geometry families despite unchanged image-formation and localization physics (Table 3; Figs. 4, 5, and 7). In simple-cyst-like exact geometry, the 0.52 formula had median APE 0.72% with no >15% failures, whereas 0.2 π and 0.625 had median APE 19.96% and 19.32%. At U, 0.52 remained comparatively accurate, with median APE 6.26% and 0% operational failure, compared with median APE 13.26% and 15.35% for the comparator formulas.

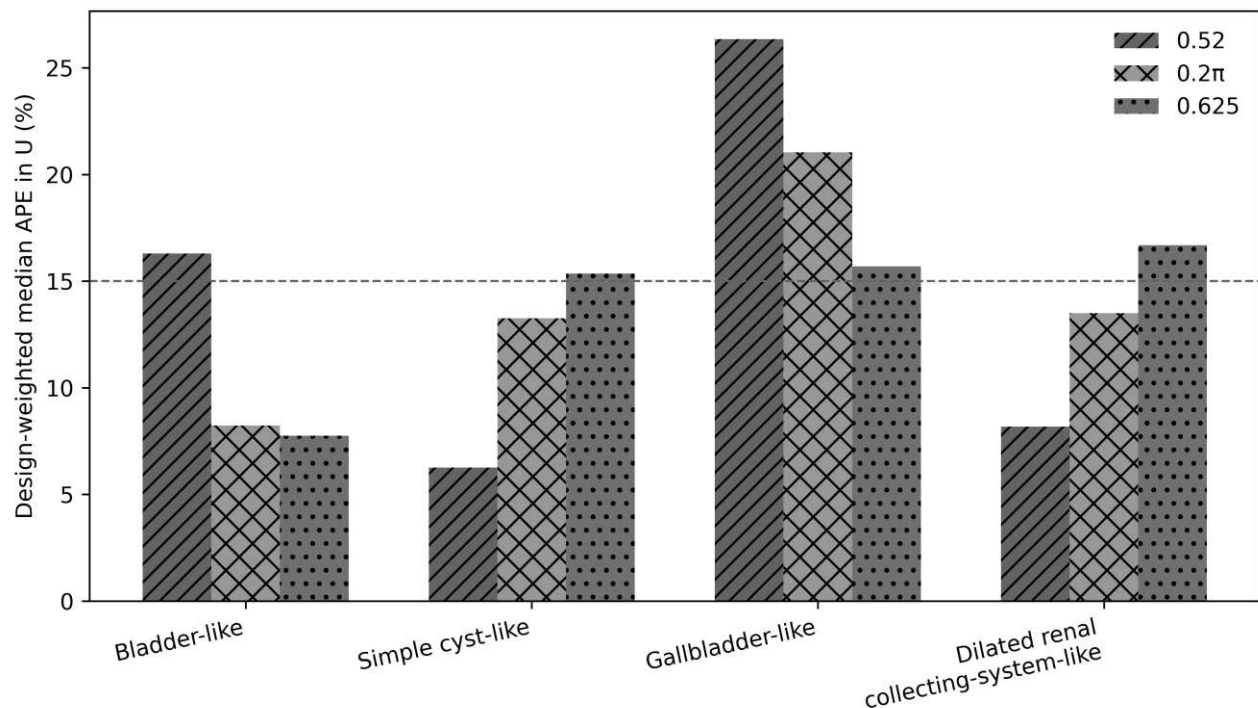


Fig. 4 Cross-geometry behavior of design-weighted U-stage median APE. The bladder-like geometry is the quantitative primary experiment; the three additional anatomy-inspired cavity families are a prospective transferability extension and are not prevalence samples

Gallbladder-like geometry showed substantially greater failure. Median APE with 0.52 progressed from 7.40% at G to 14.27% at A and 26.35% at U. At U, median APE was 21.04% for 0.2π and 15.70% for 0.625. Operational >15% failure was 62.9%, 52.2%, and 59.7%. All 16 U acquisitions were measurable and passed technical QC, showing that high volumetric error persisted despite technically acceptable interface localization.

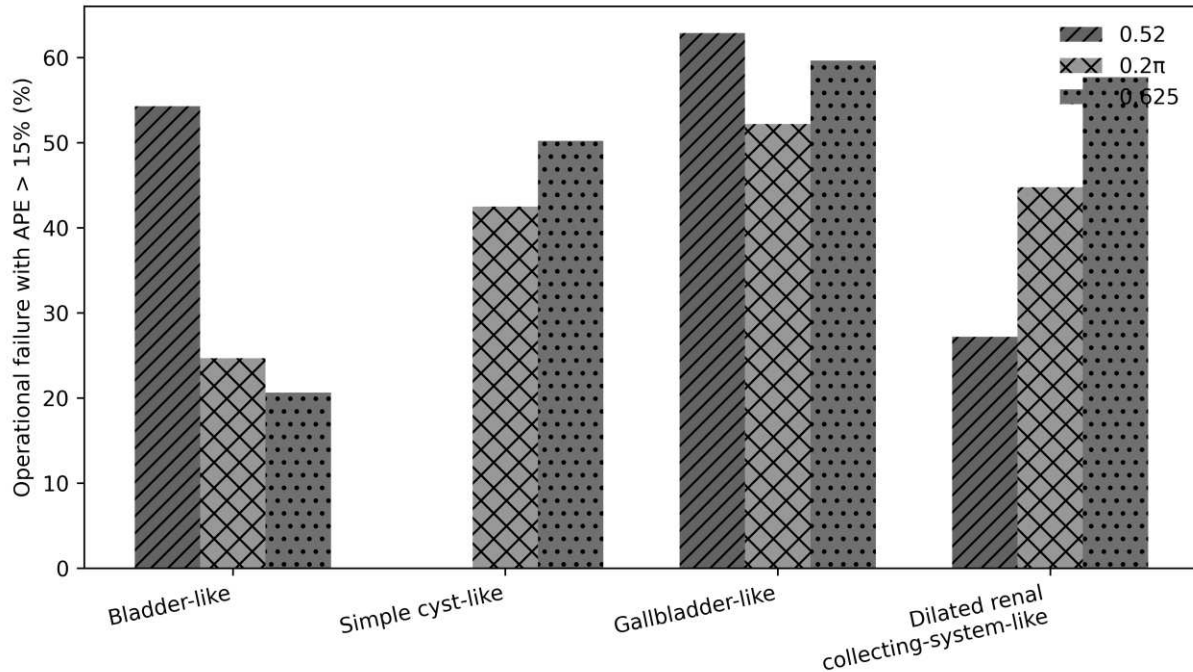


Fig. 5 Cross-geometry operational failure with APE >15% at the U stage. Extension classes had no FOV-unimageable acquisitions, whereas the primary bladder-like U sample retained five prespecified FOV-unimageable acquisitions

In the dilated renal collecting-system-like family, the 0.52 formula had median APE 2.89% at G, 3.98% at A, and 8.18% at U. U-stage operational >15% failure was 27.2% for 0.52, 44.7% for 0.2 π , and 57.7% for 0.625.

No formula was uniformly optimal. In the primary bladder-like U layer, the 0.625 formulation had the smallest APE in 45.8% of design-weighted measurable acquisitions. In contrast, 0.52 was the lowest-error formula in 90.2% of simple-cyst-like U acquisitions and 72.9% of renal collecting-system-like U acquisitions. In gallbladder-like U acquisitions, 0.625 most often had the lowest APE (45.5%), but its >15% failure fraction remained 59.7%. These results support morphology-dependent failure of fixed 2D volumetric approximations rather than a universal formula ranking.

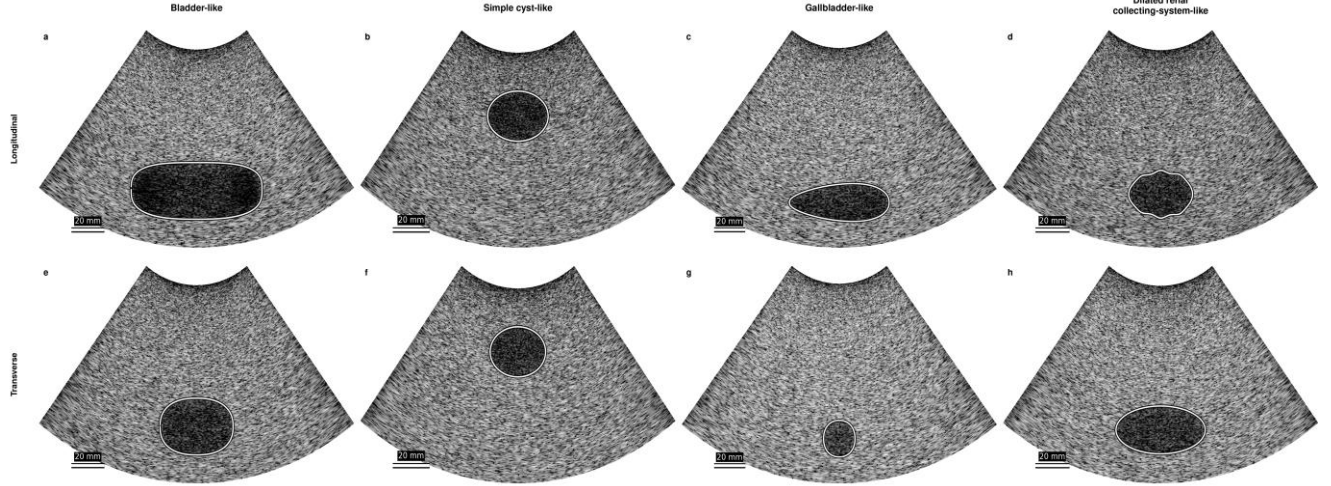


Fig. 7 Representative synthetic B-mode acquisitions from the prospectively selected virtual imaging trial. Columns show bladder-like, simple cyst-like, gallbladder-like, and dilated renal collecting-system-like geometry families; rows show longitudinal (a-d) and transverse (e-h) acquisition planes. Cases were selected without reference to volumetric error: within each geometry family, an imageable ideal-stratum acquisition nearest the median true volume of the frozen population was used. The superimposed black-and-white contour denotes the exact fluid-cavity boundary and is displayed only as a visualization reference; it is not an autonomous segmentation output. Scale bars indicate 20 mm

Table 2 Primary bladder-like volumetric performance across G, A, and U

Stage	Formula	Median APE (%)	Mean PE (%)	MAE (mL)	Failure >15% (%)	Failure >10% (%)
G	0.52	12.57	-11.12	15.65	38.9	61.3
G	0.2π	7.74	+7.39	12.77	24.9	39.3
G	0.625	7.75	+8.59	13.02	26.5	40.9
A	0.52	14.49	-13.94	18.27	47.5	68.9
A	0.2π	7.50	+3.99	11.56	19.7	37.2
A	0.625	6.97	+4.98	10.96	17.6	34.4
U	0.52	16.29	-15.53	18.07	54.3	73.1
U	0.2π	8.23	+2.06	11.03	24.7	44.6
U	0.625	7.75	+2.99	9.83	20.6	40.6

Note. G and A failure percentages are threshold exceedance among fully measurable designs. U failure percentages are operational and additionally count the five prespecified single-frame FOV-unimageable acquisitions as failures.

Table 3 Design-weighted U-stage transferability results by geometry family

Geometry family	Formula	Median APE (%)	Mean PE (%)	Failure >15% (%)	Lowest-APE fraction (%)
Bladder-like	0.52	16.29	-15.53	54.3	26.5
Bladder-like	0.2π	8.23	+2.06	24.7	27.7
Bladder-like	0.625	7.75	+2.99	20.6	45.8
Simple cyst-like	0.52	6.26	-5.37	0.0	90.2
Simple cyst-like	0.2π	13.26	+14.34	42.5	9.8
Simple cyst-like	0.625	15.35	+14.93	50.2	0.0
Gallbladder-like	0.52	26.35	-20.69	62.9	21.6
Gallbladder-like	0.2π	21.04	-4.17	52.2	32.9
Gallbladder-like	0.625	15.70	-2.15	59.7	45.5
Dilated renal collecting-system-like	0.52	8.18	-3.39	27.2	72.9
Dilated renal collecting-system-like	0.2π	13.50	+16.74	44.7	23.0
Dilated renal collecting-system-like	0.625	16.68	+18.20	57.7	4.2

Extension percentages are controlled-design transferability results and are not estimates of organ-specific clinical prevalence.

Discussion

This virtual imaging trial separates mechanisms that conventional ultrasound volumetry usually combines. Substantial error was already present with exact objects and ideal orthogonal planes, showing that part of the error arises from the geometry/formula itself. Plane perturbation then moved error in either direction, and synthetic image formation shifted mean dimensions by less than 1 mm while still changing volumetric error. A distinct FOV mechanism made some large bladder-like acquisitions unimageable in a single frame. The practical limit of 2D volumetry cannot therefore be explained by a single coefficient or by image-processing error alone.

The effective-coefficient analysis explains this pattern. For formulas based on $L \times W \times D_T$, the estimated-to-true volume ratio is $\frac{c}{k_{true}}$, where c is the selected fixed coefficient. Changing c can move the center of the error distribution, but it cannot eliminate case-to-case variation in k_{true} . The identical log-error dispersion of the 0.52 and 0.2π formulations at every stage is the expected consequence of that relationship. In the bladder-like G population, k_{true} was most strongly associated with longitudinal and transverse shape exponents rather than absolute volume. This supports the hypothesis that morphology materially constrains volumetric validity and clarifies why coefficient retuning alone cannot make a single formula universal.

These acquisition results are consistent with broader quantitative-ultrasound evidence. Ritsche et al. reported excellent reliability for 3D ultrasound muscle-volume measurements despite systematic differences from MRI; agreement depended on muscle shape and location and improved after the acquisition protocol was modified [1]. That comparison illustrates the distinction central to this study: repeatability and technically adequate images do not by themselves establish volumetric validity. Simulation work in shear-wave ultrasound has likewise used controlled models to isolate measurement confounders that cannot be readily separated in vivo [15]. Here, gallbladder-like U acquisitions passed localization QC but retained high volumetric error because a fixed ellipsoid-like approximation did not represent their morphology well.

The transferability extension shows how strongly performance depends on morphology. The simple-cyst-like family was intentionally close to a triaxial ellipsoid, for which the theoretical coefficient $\frac{\pi}{6} \approx 0.524$ is close to 0.52; correspondingly, the 0.52 formula was nearly exact in G and remained the strongest of the three formulas at U. This is consistent with the smooth, round-to-oval morphology described for simple renal cysts [7, 8] and with clinical use of ellipsoid volume approximations [9]. Gallbladder-like taper and pearness broadened k_{true} and produced high error across all three formulas, despite the continued clinical use of ellipsoid methods for gallbladder volume [10, 11]. The renal collecting-system-like family introduced a different non-ellipsoidal topology through funneling and communicating lobulation, again changing the relative formula ranking [12, 13]. The practical point is that every fixed coefficient encodes a shape model, and that model becomes less adequate as morphology changes.

FOV failure is a separate constraint because changing the formula cannot correct it. All five bladder-like U acquisitions that were prospectively unimageable occurred in the largest volume quartile. This does not establish a universal 188-mL clinical threshold; the boundary is specific to the frozen probe geometry, depth, and single-frame acquisition rule. It shows that single-frame 2D volumetry can fail before caliper placement when the target does not fit within the available view. Multi-sweep, 3D, or alternative probe strategies address a different problem from coefficient calibration.

Strengths and limitations

The study design limited several sources of retrospective bias. True cavity volume and exact plane intersections were known. The geometry and acquisition spaces were frozen before outcome generation, G and A used the full primary design, and the prospectively stratified U subset retained design weights. FOV-unimageable cases were not replaced, and the three wall-localization QC exceptions remained in the primary analysis with explicit sensitivity testing. Cross-simulator and curved-interface checks preceded definitive production, but the scientific outcomes came from one locked high-throughput pipeline. The multistructure extension was reduced prospectively, before any class-specific U outcome was inspected.

The synthetic design also sets clear limits on interpretation. The geometry families are synthetic and omit pathology, wall thickness, deformation, respiratory or other physiologic motion, and many tissue-specific acoustic effects. The target-aware localizer intentionally isolates image-formation and local interface error and is not a model of an independent clinical observer. The definitive U layer used one convex-probe configuration and one acquisition

protocol; different transducers, frequencies, compounding strategies, or multi-frame techniques could change imageability and localization error. The prespecified 15% primary and 10% secondary APE thresholds are comparative tolerances for this virtual trial rather than universally validated clinical cutoffs. Only 16 definitive U acquisitions per extension family were retained, so class-specific percentages demonstrate transferability and heterogeneity rather than precise organ-specific rates. Direct Sobol first-order and total-order indices could not be recovered from the archived single scrambled-Sobol space-filling design without a different structured sampling design or an additional surrogate analysis; morphology relationships were therefore reported descriptively rather than as formal variance decompositions.

These results suggest that 2D ultrasound volume should be treated as a model-based measurement rather than a direct geometric truth. The same measurement problem arises when serial ultrasound volumes are interpreted in settings such as simple renal cyst follow-up, gallbladder volume assessment, and dilated collecting-system evaluation [9-13]. When accuracy requirements are stringent, a fixed coefficient can correct average bias only within the morphology range for which its implicit shape assumption is appropriate. Greater robustness will likely require acquisition protocols that capture true maximal planes, morphology-aware calibration, or three-dimensional measurements when anatomy or FOV makes the ellipsoid assumption unreliable.

Conclusion

Fixed-coefficient 2D ultrasound volumetry can exceed prespecified error limits before image formation because each formula imposes an intrinsic shape model. Acquisition-plane selection can worsen or partially compensate that error, while image formation adds a smaller but measurable contribution and single-frame field of view creates a separate failure mechanism. Formula performance changed across the controlled cavity morphologies, and no evaluated coefficient was uniformly optimal. Two-dimensional volumetry should be interpreted as a model-based measurement whose validity depends on morphology and acquisition geometry, not as a direct geometric truth.

References

1. Ritsche P, Ledergerber R, Pansini M, Santini F, Faude O (2026) Assessment of lower limb muscle volume using 3D ultrasonography: validity and reliability compared to MRI. *J Imaging Inform Med* 39:1507-1518. <https://doi.org/10.1007/s10278-025-01624-1>
2. Abadi E, Segars WP, Tsui BMW et al (2020) Virtual clinical trials in medical imaging: a review. *J Med Imaging (Bellingham)* 7:042805. <https://doi.org/10.1117/1.JMI.7.4.042805>
3. Bernardino G, Garcia D (2024) PyMUST: an open-source Python library for the simulation and analysis of ultrasound. In: 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS). <https://doi.org/10.1109/UFFC-JS60046.2024.10793881>
4. Garcia D (2022) SIMUS: an open-source simulator for medical ultrasound imaging. Part I: theory & examples. *Comput Methods Programs Biomed* 218:106726. <https://doi.org/10.1016/j.cmpb.2022.106726>
5. Cigier A, Varray F, Garcia D (2022) SIMUS: an open-source simulator for medical ultrasound imaging. Part II: comparison with four simulators. *Comput Methods Programs Biomed* 220:106774. <https://doi.org/10.1016/j.cmpb.2022.106774>
6. Treeby BE, Jaros J, Rendell AP, Cox BT (2012) Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using a k-space pseudospectral method. *J Acoust Soc Am* 131:4324-4336. <https://doi.org/10.1121/1.4712021>
7. Simms RJ, Ong ACM (2014) How simple are 'simple renal cysts'? *Nephrol Dial Transplant* 29(Suppl 4):iv106-iv112. <https://doi.org/10.1093/ndt/gfu106>
8. Silverman SG, Pedrosa I, Ellis JH et al (2019) Bosniak classification of cystic renal masses, version 2019: an update proposal and needs assessment. *Radiology* 292:475-488. <https://doi.org/10.1148/radiol.2019182646>
9. Karakas HM, Yildirim G (2023) Percutaneous treatment of simple renal cysts with 24-h-interval ethanol sclerotherapy. *Radiol Bras* 56:8-12. <https://doi.org/10.1590/0100-3984.2022.0041>
10. Dodds WJ, Groh WJ, Darweesh RM, Lawson TL, Kishk SM, Kern MK (1985) Sonographic measurement of gallbladder volume. *AJR Am J Roentgenol* 145:1009-1011. <https://doi.org/10.2214/ajr.145.5.1009>

11. Lucius C, Braden B, Jenssen C et al (2025) Ultrasound of the gallbladder-an update on measurements, reference values, variants and frequent pathologies: a scoping review. *Life (Basel)* 15:941. <https://doi.org/10.3390/life15060941>
12. Koff SA, Binkovitz L, Coley B, Jayanthi VR (2005) Renal pelvis volume during diuresis in children with hydronephrosis: implications for diagnosing obstruction with diuretic renography. *J Urol* 174:303-307. <https://doi.org/10.1097/01.ju.0000161217.47446.0b>
13. Esser M, Tsiflikas I, Jago JR et al (2025) Semiautomatic three-dimensional ultrasound renal volume segmentation in pediatric hydronephrosis: interrater agreement and correlation to conventional hydronephrosis grading. *Pediatr Radiol* 55:1298-1307. <https://doi.org/10.1007/s00247-025-06249-8>
14. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307-310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
15. Miller EJ, Jin YM, Jiang J (2026) Investigating confounding factors in shear wave speed measurements of fibrotic liver tissues: a computer simulation study. *Phys Eng Sci Med* 49:959-969. <https://doi.org/10.1007/s13246-026-01712-8>
16. Oliveira WDS (2026) Supporting data and code for Geometric, acquisition and image-formation limits of two-dimensional ultrasound volumetry across simulated fluid-filled structures. Zenodo. <https://doi.org/10.5281/zenodo.22258803>

Statements and Declarations

Funding

No funds, grants, or other support were received for this study or the preparation of this manuscript.

Competing interests

The author has no relevant financial or non-financial interests to disclose.

Author contributions

Wesley da Silva Oliveira was responsible for conceptualization, methodology, software, formal analysis, investigation, data curation, visualization, manuscript drafting, review and editing, and final approval.

Ethics approval

Not applicable. This fully computational study used only synthetic geometries and simulated imaging data and involved no human participants, animals, human or animal data, or biological material.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Data availability

The definitive datasets, synthetic B-mode ultrasound data, scientific code, analysis outputs, and reproducibility documentation supporting the findings of this study are publicly available in Zenodo [16] (<https://doi.org/10.5281/zenodo.22258803>).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.pdf](#)