Performance Evaluation of Ultrasound Systems: Advanced Methods and Applications

NJ Hangiandreou PhD, Z Long PhD, DJ Tradup RDMS,
SF Stekel BS, A Ferrero PhD, AT Tao PhD,
W Zhou PhD, JE Browne PhD*, N Strissel RDMS, RVT

Mayo Clinic
Rochester, Minnesota
*Dublin Institute of Technology

MEDPHYS 3.0

The need for advanced testing methods, and translation to routine clinical physics practice

- Traditional ultrasound performance assessment

Real-time *B*-mode ultrasound quality control test procedures[a)]
Report of AAPM Ultrasound Task Group No. 1

Mitchell M. Goodsitt[b)] and Paul L. Carson
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109-0030

Scott Witt[c)]
Gammex RMI, Middleton, Wisconsin 53562-0327

David L. Hykes
Department of Radiation Therapy, St. Joseph Hospital and Health Center, Lorain, Ohio 44052

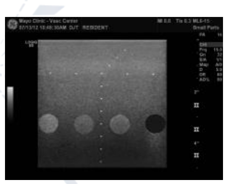
James M. Kofler, Jr.
Department of Radiology, Mayo Clinic, Rochester, Minnesota 55904

(Received 30 March 1998; accepted for publication 3 June 1998)

Med. Phys. 25 (8), August 1998

MEDPHYS 3.0

- Traditional ultrasound performance measurements
  - Measurements focus on individual, specific measures of image quality obtained from a standard quality control phantom



MEDPHYS 3.0

- We have used traditional performance assessment approaches for ultrasound system purchase decisions and acceptance testing, and have successfully added value to the selection process
- *However*, we have encountered limitations:
  - Difficulties relating these measurements to sonographer and radiologist impressions of performance
  - Clinical performance assessment of new scan modes
  - Challenges optimizing Doppler in clinical Doppler capabilities
  - Maximizing value in ultrasound quality assurance programs

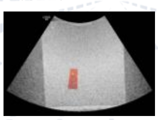
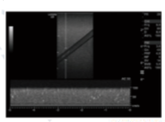
The rest of the talk:

Case studies involving performance assessment in three areas that push us beyond traditional methods

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

(2) Shear wave elastography imaging performance assessment

(3) Spectral and color Doppler performance optimization

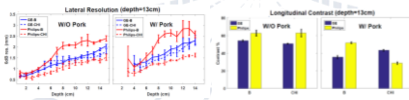




(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Clinical problem: We have added an ultrasound scanner from a 2nd vendor to an existing US-guided RF ablation practice
  - Pre-ablation ultrasound exam and (usually) next day ablation procedure of (mainly) liver lesions
  - Radiologists: “Will new system match B-mode performance level of existing systems, so they can be used interchangeably?”

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

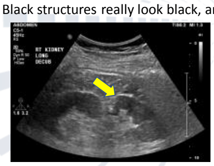
- Methods
  - Define clinical task: Visualize anechoic targets for intervention *(not search/detection)*, 1-2 cm diameter, at depths of 6-13 cm
  - CIRS 040 phantom, *with and without a layer of pork belly tissue*
  - Spatial resolution (UltraiQ, Cablon)
  - Anechoic target contrast (custom Matlab SW)
  - Sample side-by-side clinical images

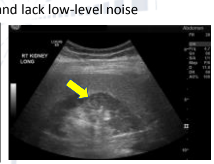


- *Some differences were seen, but none judged to be clinically important…*

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

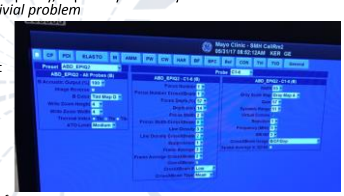
- **New clinical problem:** "You say they are ~equivalent but we really like the images on the new system – *Can you make them appear the same?*"
  - Higher contrast makes the lesions really pop
  - Tissue planes and the edges of lesions and other structures are nice and sharp
  - Black structures really look black, and lack low-level noise





(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Matching or optimizing image quality, especially across systems from different vendors, is a *complex, non-trivial problem*
  - There may not be a 1:1 correspondence between scan controls, and control calibrations may be different
  - Body parts ("models")
  - Imaging presets
  - US probes
  - Imaging modes
  - Scanner SW version



Need to address potentially many combinations of these factors…

MEDPHYS 3.0

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Matching or optimizing image quality is a problem not *usually* addressed to clinical physicists
  - We're not excluded, just not sought out
- Time-honored "clinical" approach:
  - Vendor applications specialist works with practice sonographers to establish initial presets (scan patients for 1-5 days, tweaking presets)
  - Practice sonographers scan patients, tweak settings, review results, change presets
  - Iterate for months … years
  - Buy new scanner and begin process again…

MEDPHYS 3.0

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Challenges of the "clinical" approach:
  - Presents an enormous sampling challenge: Pt habitus x Finding x Sonographer x Radiologist
    - 8-10 exams/room/day → slow process (hence perpetual iteration)
    - Not very systematic due to random presentation of patient habitus and findings
  - Image quality in any particular exam is a function of the (unpredictable) "scan-ability" of the patient and the sonographer skills and experience
  - Clinical feedback from sonographers and radiologists is generally unstructured, usually based on a single image, and can be passionate and "viral"

MEDPHYS 3.0

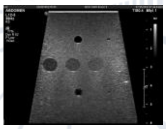
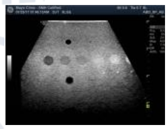
(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Goal of "physics" approach to image quality optimization: Simulate a variety of patient habitus and findings and use side-by-side imaging to systematically define a reasonable scan parameter "neighborhood" *in which to begin the traditional clinical process*

MEDPHYS 3.0

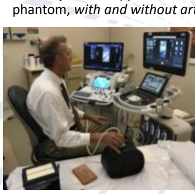
(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Initial, **8-step** “physics” approach to image quality optimization:
  1. Definition of clinical task (feedback from initial radiologist review)
  2. Analysis of scanner controls that will likely come into play
     (model, map, DR, SRI, rejection, suppression, spatial compounding // frequency, gain, focus, …)
  3. Adjust model, map, and DR w traditional QC phantom

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Initial physics approach to image quality optimization:
  4. Adjustment of all controls, including proprietary image processing and noise rejection/suppression, scanning an anthropomorphic abdomen phantom, *with and without artificial distortion layers*

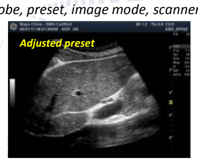



(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Initial physics approach to image quality optimization:
  5. Continued adjustment of all controls while scanning a volunteer, *with and without artificial distortion layers*

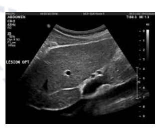

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Initial physics approach to image quality optimization:
  6. Review changes with sonographers and request patient comparison images
     - Obtain image comparison feedback from radiologists and other sonographers
  7. Assess image similarity using traditional performance measures (QC phantom and UltraiQ)
  8. *REPEAT for all combinations of probe, preset, image mode, scanner SW version…*



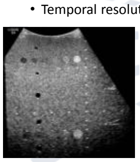
(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

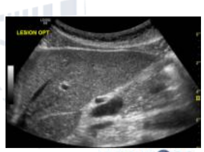
- **Possible improvements to the initial physics approach:**
  - Additional “traditional” measurements
  - Validated set of phantom distortion layers
  - Widely available target-SNR measurement tools
  - Improved methods and tools for obtaining feedback from clinical users
  - Ultrasound practice analytics data



(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Possible improvements to the initial physics approach:
  Additional “traditional” measurements
  - Speckle texture
  - Temporal resolution





(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Possible improvements to the initial physics approach:
  Validated set of phantom distortion layers
  - Include all sources of distortion (reverb, haze, phase aberration, attenuation)
  - Mimic *wide range* of patient body habitus seen in the practice
  - Use with ~any other phantom (diagnostic or therapeutic), and volunteers

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Possible improvements to the initial physics approach:
  Widely available target-SNR measurement tools

Some nice target phantoms are commercially available…



…but other targets are needed

MEDPHYS 3.0

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Possible improvements to the initial physics approach:
  Widely available target-SNR measurement tools

IMPROVED METHOD FOR DETERMINING RESOLUTION ZONES IN ULTRASOUND PHANTOMS WITH SPHERICAL SIMULATED LESIONS

James M. Kofler, Jr.* and Ernest L. Madsen†

*Department of Radiology, Mayo Clinic, Rochester, MN, USA; and †Department of Medical Physics, University of Wisconsin, Madison, WI, USA

Ultrasound in Med. & Biol., Vol. 27, No. 12, pp. 1667–1676, 2001

The Resolution Integral – a tool for characterising the performance of diagnostic ultrasound scanners

Carmel M Moran[1], Scott Inglis[2] and Stephen D Pye[2]

[1]Medical Physics, Centre for Cardiovascular Science, Queen's Medical Research Inst, University of Edinburgh, Ec
[2]Medical Physics, Royal Infirmary of Edinburgh, Edinburgh, UK

Ultrasound 2014; 22: 37–43. DOI: 10.1177/1742271X13518202

…and commercial or open-source analysis software is lacking

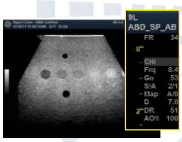
MEDPHYS 3.0

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Possible improvements to the initial physics approach:
  Improved methods and tools for obtaining feedback from clinical users (and improved understanding this feedback)
  - Email link to network app, preloaded with image pairs for evaluation
  - Obtain ratings data for ~20 image pairs in <15 min
  - Specific questions and terminology must be carefully considered
    - Which image is preferred, A or B?
    - Rate image B compared with image A in terms of its clinical utility:
      - Superior
      - Equivalent
      - Inferior but clinical acceptable
      - Not clinically acceptable

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Possible improvements to the initial physics approach:
  Ultrasound practice analytics data
  - Gather and database information from US images, public and private DICOM headers, and DICOM structured report (SR) messages



SOPClassUID	US Image
SOPInstanceUID	[illegible]
StudyInstanceUID	1.2.840.113717.2.21298431.1
StudyDateTime	5/12/2017 3:54
StudyID	21298431-1
StudyDescription	US Abdomen Limited
SeriesInstanceUID	[illegible]
SeriesDateTime	5/12/2017 3:54
SeriesNumber	1
InstanceNumber	18
ContentDateTime	5/12/2017 4:22
AccessionNumber	21298431-1
Modality	US
Manufacturer	GE Healthcare
InstitutionName	Mayo Clinic - SMH Call Rm
StationName	LE97QM
OperatorName	DJD
ManufacturerModelName	LOGIQE9
SoftwareVersion	LOGIQE9:R3.1.3

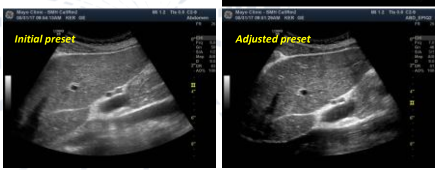
Subject Age	Patient Characteristics	Vascular Ultrasound Procedure Report	
Prox AP	Aorta	Findings (Finding Site = General Vascular, Image Mode = 2D mode)	Vascular Ultrasound Procedure Report
Distal AP	Aorta	Findings (Finding Site = General Vascular, Image Mode = 2D mode)	Vascular Ultrasound Procedure Report
Gall Bladder Wall	Gall Bladder	Findings (Finding Site = Artery of Abdomen, Image Mode = 2D mode)	Vascular Ultrasound Procedure Report
CHD	Biliary Ducts	Findings (Finding Site = General Vascular, Image Mode = 2D mode)	Vascular Ultrasound Procedure Report
CBD	Biliary Ducts	Findings (Finding Site = General Vascular, Image Mode = 2D mode)	Vascular Ultrasound Procedure Report

(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

- Possible improvements to the initial physics approach:
  Ultrasound practice analytics data
  - US image analytics data can facilitate several steps in this process…
    - Use DICOM SR as a guide to define the clinical imaging task being studied, e.g. lesion dimensions, depths, echogenicity
    - Analyze the actual usage rates of presets, probes, and image modes in a large group of prior exams involved in the clinical imaging task
    - After implementation of a new preset, observe the usage rates of the new and previous presets, and identify sonographer patterns of use
    - Analyze the image-to-image control changes actually made by sonographers in a large group of exams, edit the presets to optimize

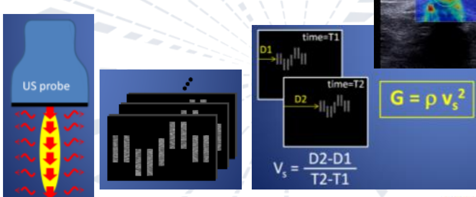
(1) Performance standardization of two ultrasound systems for US-guided RF ablation practice

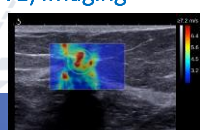
- New ablation preset for 3 probes (C1-6, C2-9, 9L), and fundamental and harmonic imaging modes
- Good *initial* clinical feedback…



(2) Shear wave elastography (SWE) imaging performance assessment

- High-level ultrasound SWE imaging process





(2) Shear wave elastography imaging performance assessment

- Two types of clinical applications
  - Assessment of liver stiffness to diagnose diffuse disease
  - Assessment of stiffness of focal targets

- Recent clinical SWE performance testing requests
  1. Acceptance test liver SWE capability of 23 upgraded clinical scanners
  2. Compare median nerve SWE measurements from 2 scanners

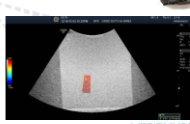
(2) Shear wave elastography imaging performance assessment

- RSNA QIBA efforts: "…to unite researchers, healthcare professionals, and the industry to advance *quantitative imaging* and the use of imaging biomarkers in clinical trials and clinical practice."
  → Ultrasound shear wave speed
  - **Development of visco-elastic tissue-mimicking phantoms**
  - Multi-vendor system comparisons in both phantoms and patients

(2) Shear wave elastography imaging performance assessment

1. Acceptance test liver SWE capability of 23 upgraded clinical scanners
   (Presented by Z Long PhD at 2016 AAPM Annual Meeting)

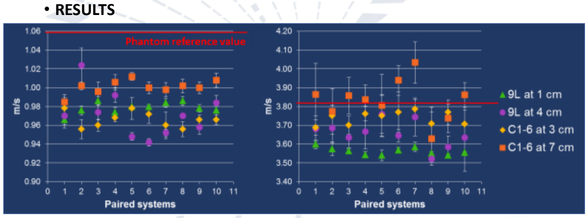
- **METHODS**
  - 10 GE LOGIQ E9 XDclear 2.0 US scanners with C1-6 and 9L transducers
  - CIRS shear wave liver fibrosis phantoms
    - Young's modulus of ~3.5 and 45 kPa
    - Two measurement depths for each probe
  - Three operators, 5 repeat datasets
  - Investigated different coupling media and phantom homogeneity



(2) Shear wave elastography imaging performance assessment

1. Acceptance test liver SWE capability of 23 upgraded clinical scanners

- **RESULTS**



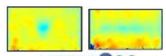
(2) Shear wave elastography imaging performance assessment

1. Acceptance test liver SWE capability of 23 upgraded clinical scanners

- **CONCLUSIONS**
  - Measurement differences between systems were within about 5% of the mean value
  - Bias towards underestimation of shear wave speed is seen, but is consistent with bias seen between systems by the QIBA group
  - No differences between gel and salt water coupling media were seen
  - Modulus variations at different phantom locations were <2.5%
  - *US systems were accepted, and questions re methods were resolved*

2. Compare median nerve SWE measurements from 2 scanners…

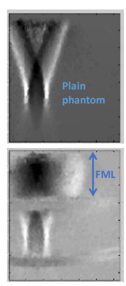
Evaluation of Two Ultrasound Systems for Shear Wave Elastography Measurements of Small Targets
Z Long*, D Tradup , S Eby , P Song , S Chen , K Glazebrook , N Hangiandreou , Mayo Clinic, Rochester, MN
Presentations
WE-DE-708-6 (Wednesday, August 2, 2017) 10:15 AM - 12:15 PM Room: 708



MEDPHYS 3.0

(2) Shear wave elastography imaging performance assessment

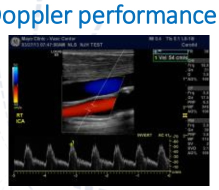
- New imaging mode + new testing tools + new methods
  → Need for increased care and skeptical, critical analysis of data
  → Need for good understanding of the physics and engineering of the new imaging mode
- *EXAMPLE:* Testing shear wave elastography using SWE liver phantoms with a fat-mimicking layer (FML) created a wave pattern that fooled the wave tracking algorithm and resulted in biased velocity measurements
  - These wave patterns *not* encountered in any patient testing
- *Partnership with vendor or research colleagues is key*



MEDPHYS 3.0

(3) Spectral and color Doppler performance optimization

- Majority of reported ultrasound performance testing work seems focused on 2D grayscale imaging
- Use of Doppler in clinical practice
  - Grayscale images = 67.2%
  - Doppler images = 32.8%
    - Duplex Doppler = 18.5%
    - Color Doppler = 14.3%



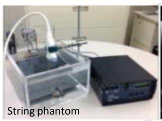
- Recent clinical Doppler performance measurement and optimization requests
  1. Standardize calibration of clinical peak velocity measurements across Doppler angles, probes, scanner models, and vendors
  2. Optimize clinical color Doppler presets, improving scanner performance for filling vessels and rejecting flash artifact (coloring of slowly moving tissue)
  3. Assess the clinical reproducibility of peak systolic velocity measurements in the carotid arteries

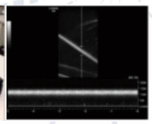
MEDPHYS 3.0

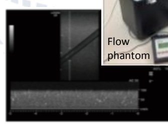
(3) Spectral and color Doppler performance optimization

1. Standardize calibration of clinical peak velocity measurements across Doppler angles, probes, scanner models, and vendors

- Correct for intrinsic spectral broadening errors
- String phantoms or flow phantoms *(or both?)*
  - Ability to mimic clinical imaging conditions
  - Accurate (and ideally, verifiable) *velocity* specification



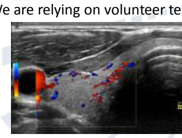


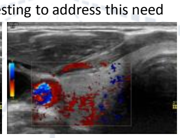


(3) Spectral and color Doppler performance optimization

2. Optimize clinical color Doppler presets, improving scanner performance for filling vessels and rejecting flash artifact (coloring of slowly moving tissue)

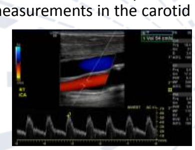
- We have used flow phantoms and common Doppler tests for pre-purchase scanner evaluations, *however they are limited in this particular application*
  - Adding a motion challenge for color Doppler assessment is similar to adding distortion layers to challenge grayscale imaging performance → Better simulation of real patient imaging
- We are relying on volunteer testing to address this need



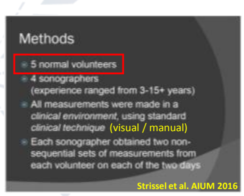


(3) Spectral and color Doppler performance optimization

3. Assess the clinical reproducibility of peak systolic velocity measurements in the carotid arteries



NASCET peak systolic velocity (PSV) thresholds for internal carotid artery (ICA) stenosis:
ICA PSV < 125 cm/s → normal to <50% stenosis
ICA PSV > 230 cm/s → >70% stenosis



MEDPHYS 3.0

(3) Spectral and color Doppler performance optimization

3. Assess the clinical reproducibility of peak systolic velocity measurements in the carotid arteries

Overall PSV standard deviation for *all measurements*: ±8 cm/s

Applications of the Repeatability of Quantitative Imaging Biomarkers: A Review of Statistical Analysis of Repeat Data Sets

Huiman X. Barnhart* and Daniel P. Barboriak†

*Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27710, USA; †Department of Radiology, Duke University Medical Center, Durham, NC 27710, USA

Translational Oncology (2009) 2, 231–235

Quantitative Imaging Biomarkers Alliance RSNA

MEDPHYS 3.0

(3) Spectral and color Doppler performance optimization

3. Assess the clinical reproducibility of peak systolic velocity measurements in the carotid arteries

Different Vessels	Standard Deviation
Distal CCA PSV	7.0 cm/s
Bulb PSV	7.1 cm/s
ICA PSV	9.6 cm/s

Sonographers	Standard Deviation
Sonographer 1*	5.7 cm/s
Sonographer 2	6.4 cm/s
Sonographer 3*	5.5 cm/s
Sonographer 4	4.6 cm/s

*More experienced sonographers

Next steps:
Repeat this experiment in a larger group of patients with carotid disease

MEDPHYS 3.0

Conclusions

- Traditional performance assessment methods for grayscale imaging are very useful for many clinical physics applications
  - Limitations and complications may be encountered when applications bridge the gap between basic system performance and impressions of clinical users
- Enhancements to the clinical physics toolkit are needed
  - Phantoms that better mimic clinical scanning conditions
  - Performance measures oriented towards target detection & characterization, and validated, commercial or open-source software analysis tools
  - More systematic and efficient approaches to gathering clinical feedback
  - Better utilization of volunteer and patient image data
  - Utilization of ultrasound practice analytics data

MEDPHYS 3.0

Conclusions

- Traditional performance assessment approaches can be productively, and carefully adapted to new scanner modes
  - Partnership with vendor or research colleagues is key
- Combinations of individual components into larger protocols may allow more complex problems to be addressed
  - Traditional QC phantoms → Target phantoms → Anthropomorphic phantoms → Volunteers → Patients

- *There is growing clinical practice demand for physics assistance, often requiring advanced performance assessment approaches to be utilized or developed = opportunity*

MEDPHYS 3.0