

Reproducibility of Dynamic MR Imaging Pelvic Measurements: A Multi-institutional Study¹

Mark E. Lockhart, MD, MPH
 Julia R. Fielding, MD
 Holly E. Richter, MD, PhD
 Linda Brubaker, MD
 Caryl G. Salomon, MD
 Wen Ye, MD
 Christiane M. Hakim, MD
 Clifford Y. Wai, MD
 Alan H. Stolpen, MD
 Anne M. Weber, MD

¹ From the Departments of Radiology (M.E.L.) and Obstetrics and Gynecology (H.E.R.), University of Alabama at Birmingham, 625 19th St South, JTN Room 363, Birmingham, AL 35249; Department of Radiology, University of North Carolina, Chapel Hill, NC (J.R.F.); Departments of Obstetrics & Gynecology and Urology (L.B.) and Radiology (C.G.S.), Loyola University, Chicago, Ill; Department of Biostatistics, University of Michigan, Ann Arbor, Mich (W.Y.); Department of Radiology, University of Pittsburgh, Pittsburgh, Pa (C.M.H.); Department of Obstetrics and Gynecology, University of Texas–Southwestern, Dallas, Tex (C.Y.W.); Department of Radiology, University of Iowa, Iowa City, Iowa (A.H.S.); Edward Hines, Jr., VA Hospital, Hines, Ill (C.G.S.); and National Institute of Child Health and Human Development, Bethesda, Md (A.M.W.). Received November 16, 2007; revision requested January 15, 2008; revision received May 19; accepted May 30; final version accepted June 16. Supported in part by National Institute of Child Health and Human Development grants 2U10 HD41249, 2U10 HD41250, 2U10 HD41261, 2U10 HD41267, 1U10 HD41248, 1U10 HD41263, 1U10 HD41268, and 1U10 HD41269. **Address correspondence** to M.E.L. (e-mail: mlockhart@uabmc.edu).

© RSNA, 2008

Purpose:

To assess the reproducibility of bone and soft-tissue pelvimetry measurements obtained from dynamic magnetic resonance (MR) imaging studies in primiparous women across multiple centers.

Materials and Methods:

All subjects prospectively gave consent for participation in this institutional review board–approved, HIPAA-compliant study. At six clinical sites, standardized dynamic pelvic 1.5-T multiplanar T2-weighted MR imaging was performed in three groups of primiparous women at 6–12 months after birth: Group 1, vaginal delivery with anal sphincter tear ($n = 93$); group 2, vaginal delivery without anal sphincter tear ($n = 79$); and group 3, cesarean delivery without labor ($n = 26$). After standardized central training, blinded readers at separate clinical sites and a blinded expert central reader measured nine bone and 10 soft-tissue pelvimetry parameters. Subsequently, three readers underwent additional standardized training, and reread 20 MR imaging studies. Measurement variability was assessed by using intraclass correlation for agreement between the clinical site and central readers. Acceptable agreement was defined as an intraclass correlation coefficient (ICC) of at least 0.7.

Results:

There was acceptable agreement (ICC range, 0.71–0.93) for eight of 19 MR imaging parameters at initial readings of 198 subjects. The remaining parameters had an ICC range of 0.13–0.66. Additional training reduced measurement variability: Twelve of 19 parameters had acceptable agreement (ICC range, 0.70–0.92). Correlations were greater for bone (ICC, ≥ 0.70 in five [initial readings] and eight of nine [rereadings] variables) than for soft-tissue measurements (ICC, ≥ 0.70 in three [initial readings] of 10 and four [rereadings] of 10 readings, respectively).

Conclusion:

Despite standardized central training, there is high variability of pelvic MR imaging measurements among readers, particularly for soft-tissue structures. Although slightly improved with additional training, measurement variability adversely affects the utility of many MR imaging measurements for multicenter pelvic floor disorder research.

© RSNA, 2008

Supplemental material: <http://radiology.rsna.org/cgi/content/full/2492072009/DC1>

Pelvic floor symptoms are common in women after childbirth (1). Objective assessment of anatomic changes and structural pathologic indicators is an important adjunct in characterization of pelvic floor symptoms resulting from childbirth. Dynamic magnetic resonance (MR) imaging is used to assess pelvic organ prolapse (2–5); however, correlation of MR findings with physical examination and cystocolpoproctography results is variable (5–7). Defecography also has high interobserver variability (8). Measurement reproducibility is important to assess, whether in the research or the clinical setting, since it may affect management of patients with pelvic floor disorders.

Previously, a few small single-center series or retrospective studies have shown variable interobserver reliability in characterizing specific anatomic findings of anal sphincter and pelvic structures demonstrated by using MR imaging (9–14). Intraobserver correlation of pelvic organ prolapse has been weak even in single-site studies (12). Moderate and moderate-to-good interobserver correlation of external anal sphincter atrophy on endoanal ($\kappa = 0.53$ – 0.56) and phased-array coil ($\kappa = 0.55$ – 0.80) MR imaging have been reported (15). Better interobserver correlation has been shown in other single-center studies (10,11). Beets-Tan et al (10) showed better interobserver correlation for the internal anal sphincter by using endoanal (intraclass correlation coefficient [ICC], 0.65) and phased-array (ICC, 0.75) MR imaging. Keller et al (11) showed high interobserver

correlation for bone measurements such as obstetric conjugate (ICC, 0.96) and interspinous distance (ICC, 0.94). However, no large multi-institutional trials have evaluated this question.

The purpose of the current study is to assess the reproducibility of bone and soft-tissue pelvimetry measurements obtained from dynamic MR imaging studies in primiparous women across several centers.

Materials and Methods

Study Design

The Childbirth and Pelvic Symptoms (CAPS) study evaluated fecal and urinary incontinence symptoms at 6 weeks and 6 months after delivery in three cohorts of primiparous subjects: Group 1, after vaginal delivery with a clinically recognized anal sphincter tear (v-tear); group 2, after vaginal delivery without a clinically recognized anal sphincter tear (vaginal control); and group 3, those who underwent cesarean delivery without labor (cesarean control) (1). The CAPS Imaging Study (CAPS-IS) (16) is a multi-institutional study in which endoanal ultrasonographic (US) and dynamic MR imaging was performed 6 months after birth in a subset of the subjects. Institutional review board approval was obtained at participating clinical sites and the Data Coordinating Center (University of North Carolina, Chapel Hill, NC). Informed consent was obtained for the imaging portion of this Health Insurance Portability and Accountability Act-compliant study. Data regarding the US findings in all subjects of this trial have been previously reported (16).

After the approval of CAPS-IS (about 12 months after CAPS began), all CAPS subjects were approached to participate in this study; 253 primiparous patients from the CAPS study agreed to participate in CAPS-IS. Of these, 247 completed the MR imaging

component of the study. We restricted analysis to MR imaging studies interpreted by both the site and the expert central readers; therefore, 47 subjects from the central reader site were not included, since there was no second reader for variability comparison, and in two cases, we did not receive both evaluations. Therefore, 198 MR images were evaluated for variability and placed in one of three groups: v-tear ($n = 93$), vaginal control ($n = 79$), and cesarean control ($n = 26$). The three groups did not significantly differ regarding age (Table 1). Each site recruited subjects in all three cohorts.

Initial MR Imaging Training and Data Acquisition

The participating radiologist from each of the six clinical sites (10–20 years experience each) attended a 1-day training session with the expert central reader (J.R.F., 20 years experience with >300 pelvic floor examinations) prior to study initiation. Training consisted of description of the desired measurements and review of measurement technique, including relevant images to be used, bone and soft-tissue land-

Advances in Knowledge

- The utility of MR imaging soft-tissue and pelvimetry parameters may be limited by high measurement variability among readers at different institutions despite standardized training.
- Interobserver agreement for bone parameters shows a trend better than for continuous soft-tissue or categoric parameters (ICC, >0.70 in eight of nine versus four of ten variables, respectively; $P = .057$).

Implication for Patient Care

- High variability in pelvic floor MR imaging measurements may limit generalizability of results.

Published online before print
10.1148/radiol.2492072009

Radiology 2008; 249:534–540

Abbreviations:

CAPS = Childbirth and Pelvic Symptoms
CAPS-IS = CAPS Imaging Study
ICC = intraclass correlation coefficient
PCL = pubococcygeal line

Author contributions:

Guarantor of integrity of entire study, M.E.L.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, M.E.L., J.R.F.; clinical studies, J.R.F., H.E.R., L.B., C.G.S., C.M.H., A.H.S., M.E.L., C.M.H., A.M.W.; statistical analysis, W.Y.; and manuscript editing, M.E.L., J.R.F., H.E.R., C.G.S., W.Y., A.H.S.

Funding:

This research was funded by the National Institute of Child Health and Human Development (grants 2U10 HD41249, 2U10 HD41250, 2U10 HD41261, 2U10 HD41267, 1U10 HD41248, 1U10 HD41263, 1U10 HD41268, 1U10 HD41269).

Authors stated no financial relationship to disclose.

marks, and use of measurement tools (Advantage Workstation; GE Healthcare, Milwaukee, Wis) by using electronic presentation consisting of multiple examples of normal and abnormal findings. Demonstration of the measurement technique with a workstation was performed. In addition, participants viewed the acquisition of a pelvic MR examination by using the standardized protocol to image a volunteer subject.

Study MR examinations were begun after test data sets submitted by each site were reviewed and approved for quality and protocol adherence by the expert central reader. Data collection was from September 2003 to March 2005. The reference standard was the results of the expert central reader. All readers remained blinded to patient group and clinical information throughout the study.

MR Imaging Technique

After subjects were instructed to void bladder and bowel, approximately 60 mL of inert US gel was placed in the rectum with the patient in the lateral decubitus position. The patient's position was changed to supine and a pelvic phased-array coil was placed around the lower pelvis. MR imaging was performed by using a standardized protocol with 1.5-T imagers. Four types of magnets were used (Symphony or Vision, Siemens Medical Systems, Erlangen, Germany; or Sigma or Echospeed, GE Healthcare). The protocol consisted of localizer images, sagittal ultrafast T2-weighted images (repetition time msec/echo time msec, 4400/90; field of view, 300; section thickness, 10 mm; matrix, 128 × 256; and number of signals acquired, one) at rest and at strain, transverse and coronal T2-weighted images (5000/132; field of view, 200; section thickness, 3 mm; matrix, 270 × 256; and number of signals acquired, two) at rest, and oblique coronal T2-weighted images (4400/90; field of view, 250; section thickness, 5 mm; matrix, 128 × 256; and number of signals acquired, one) parallel to the sacrum. The number of sections varied with patient size to include the region of interest. Images were not angled with the pelvic floor.

No intravenous or vaginal contrast agents were used. There was no bowel preparation required prior to the examination. Total imaging time was approximately 25 minutes.

Standardized measurements were made with electronic calipers at a workstation (Advantage Windows with Centricity, GE Healthcare; or IMPAX, Agfa, Peissenberg, Germany) and recorded on standardized forms. The MR examination was stripped of all measurements and protected health information, recorded on a compact disc with its appropriate research number as the sole method of identification, and sent to a central site for a second interpretation by the expert reviewer.

MR Imaging Retraining and Remeasurement of Subset Data

To assess whether additional measurement standardization training would improve interobserver variability, three radiologists (M.E.L., C.G.S., and C.M.H.) at different sites volunteered to reinterpret a mixed subgroup of MR data sets from the initial study, which had high interobserver variability (ICC, <0.70). Approximately 18 months after the initial training session, each reproducibility reader underwent an additional 6 hours of interpretive training performed by the expert central reader, in conjunction with the project statistician. Specific pelvic MR measurements were reviewed and subsequently practiced in three complete pelvic MR data sets until satisfactory interobserver agreement had been achieved, as qualitatively determined by the project statistician. The three radiologist reproducibility readers remained blinded to

clinical cohort information and initial protocol outcomes.

Following completion of training, 20 MR imaging studies were randomly selected by the Data Coordinating Center in compact disk format, devoid of personal health information, and were sent for independent reinterpretation by the three readers by using new study identifiers. The subset readings occurred over a 3-month period (18–21 months after the initial readings). Measurements were made by using software (Efilm Lite, version 1.8.2; Stentor, Foster City, Calif) embedded in each disc, allowing digital caliper and angle measurements.

MR Interpretation Parameters

Thirty individual pelvic MR measurements were obtained by two readers, one each from the site reader and the expert central reader for each patient during the initial trial (Appendix E1 [<http://radiology.rsna.org/cgi/content/full/2492072009/DC1>]). Of 30 measurement parameters, 22 were continuous and eight were categorical. Two of the variables are measurements of right and left minimal gap distance on sides where a gap at the sling insertion to symphysis was present. Since only a small portion of the readings had a gap recognized as present, these two measurements were not available in many readings and were therefore omitted in the analysis. Additionally, because of inconsistencies in the definitions used by the readers, three continuous variables (distance from bladder neck to pubococcygeal line [PCL] with strain, angle of levator plate with PCL at rest and with strain) were omitted from this analysis, mainly owing to inconsistent use of

Table 1

Patient Demographics			
Group	Mean Age (y)*	Age Range (y)	No. of Patients
Entire population	26.6 ± 6.2	15–43	198
Vaginal delivery with clinically diagnosed anal sphincter disruption	27.2 ± 6.4	15–43	93
Vaginal control	25.8 ± 5.5	18–39	79
Cesarean control	27.3 ± 7.4	16–41	26

* Data are the mean ± standard deviation.

positive and negative signs. The difference between measurements at rest and with strain was calculated for both H-line and M-line (Appendix E1 [<http://radiology.rsnajnl.org/cgi/content/full/2492072009/DC1>]), creating two new variables. Therefore, the analyses on initial readings include 19 continuous variables (nine bone and 10 soft-tissue dimensions) and eight categorical variables.

For the subset of 20 reread MR images, 25 continuous and six categorical and binary MR measurements were collected by three readers. Anal sphincter tear evaluation was not reexamined because of poor results in the initial trial; there were no agreements between the two readers and very few cases were identified as having tears by either reader. Levator symphysis gap was defined differently during the second round of training after discussion among the readers,

which prohibited comparison between the two sets of readings. Six new continuous variables were proposed and evaluated for exploratory purposes, including right and left minimal gap distances if a levator symphysis gap was present, urethral angle at rest and with strain, and vaginal length at rest and with strain. We compared 19 common continuous variables and six common categorical variables between the initial readings and the rereadings.

Statistical Analysis

ICC was calculated for each parameter (17) and is defined as the ratio of variance between images to total variance. If the readers identify the same landmarks, then the measurements should be very similar; therefore, a relatively high threshold level was selected to define acceptable reliability: $\kappa \geq 0.85$ was

considered as good reliability, $\kappa = 0.70$ to less than 0.85 was considered as acceptable reliability, and $\kappa < 0.70$ was considered as poor reliability. The ICC of the initial readings was compared with that of the rereadings. The parameters in each category (good, acceptable, poor) were compared. The ICC of the individual bone parameters and the soft-tissue parameters was also compared by using the same categories (good, acceptable, poor). The ICC results were analyzed by reader site to evaluate for any systematic differences. Of eight categorical variables in the initial trial, four were dichotomous. The others—vaginal shape, ileococcygeus contour, and two analyses of anal sphincter tears—were dichotomized for statistical analysis. Most responses regarding vaginal shape were “normal H” or “butterfly,” and most ileococcygeus muscles were categorized as “superiorly bowed.” These responses were counted as “yes” responses, minority responses were counted as “no.” For ileococcygeus muscle, “bowed superiorly” is considered as “yes,” “flat” and “bowed inferiorly” are considered as “no.” For the two anal sphincter tear questions, “cannot visualize” was also treated as “no” for statistical evaluation, although the two responses are not exactly the same.

κ Statistics were calculated (18) for all eight dichotomous variables. The six categorical variables in the rereadings were analyzed similarly. Since there were more than two readers, generalized κ (19) was calculated for rereadings. A threshold level of $\kappa \geq 0.85$ was considered as good, $\kappa = 0.70$ to less than 0.85 as acceptable, and $\kappa < 0.70$ as poor reliability.

Table 2

Difference between Site Readings and Central Expert Readings in the Initial Trial in 198 Patients

Variable	Central Expert (mm)*	Difference between Central Expert and Site Reading (mm)*	ICC
Bone tissue			
Interspinous distance	104.2 ± 8.1	3.7 ± 3.5	0.75
Intertuberous diameter	121.8 ± 9.5	-1.7 ± 5.2	0.73
Angle of pubic arch with symphysis as apex	83.5 ± 7.0	2.0 ± 5.0	0.55
Obstetric conjugate	122.3 ± 10.6	-0.8 ± 2.7	0.93
Anteroposterior outlet	111.6 ± 10.7	-1.7 ± 4.8	0.78
Sacral length	116.7 ± 14.4	1.2 ± 5.3	0.86
Sacral depth	39.7 ± 7.5	-2.5 ± 4.2	0.63
Transverse inlet	104.8 ± 7.5	7.33 ± 6.81	0.29
Transverse diameter	123.3 ± 8.2	5.2 ± 4.2	0.64
Soft tissue			
Width of levator hiatus	38.2 ± 5.5	3.8 ± 4.5	0.47
Width of right levator sling muscle	4.25 ± 1.62	0.5 ± 1.54	0.40
Width of left levator sling muscle	5.02 ± 1.95	0.93 ± 1.73	0.42
Distance from bladder neck to PCL, measured at rest	21.8 ± 5.2	-0.8 ± 2.7	0.71
H-line measured at rest	43.6 ± 7.7	5.4 ± 5.3	0.46
H-line measured during strain	52.6 ± 11.3	2.2 ± 5.8	0.77
Difference between H-line, measured at rest and during strain	9.0 ± 9.4	-3.2 ± 5.2	0.66
M-line measured at rest	12.7 ± 5.8	5.8 ± 7.7	0.13
M-line measured during strain	24.6 ± 12.4	7.5 ± 7.2	0.61
Difference between M-line, measured at rest and during strain	11.8 ± 11.4	1.8 ± 5.4	0.74

* Data are the mean ± standard deviation.

Results

Initial MR Reading

Two of 19 continuous variables had good reliability, obstetric conjugate (ICC, 0.93) and sacral length (ICC, 0.86) (Tables 2, 3). Six had acceptable reliability, including interspinous distance, intertuberous diameter, distance from bladder neck to PCL at rest, H-line, difference between M-line at rest and strain, and anteroposterior outlet. Eleven variables had poor reliabil-

ity (range, 0.13–0.66). The M-line at rest had the lowest ICC (0.13), indicating poor reliability. Among the six sites, there were parameters with high interobserver variability from every site. The variability was not limited to a subset of the readers.

Five of nine bone pelvimetry measurements and three of 10 soft-tissue measurements showed acceptable interobserver correlation on the basis of the initial training.

There was disagreement between paired readers for the eight categorical and binary variables (Table 4), particularly for the two sphincter tear measurements, with κ values of -0.023 and -0.019 . The other κ values vary from 0.12 to 0.54. The small number of enteroceles in this sample precluded adequate statistical evaluation of this parameter.

Repeat MR Readings and Outcomes

Among the 11 variables with unacceptable reliability in the initial trial, retraining improved the reliability of six variables to the acceptable level, including width of levator hiatus, angle of pubic arch, H-line with strain, difference between H-line at rest and with strain, depth of sacral hollow, transverse inlet, and transverse diameter (Table 3). Overall, seven measurements improved by at least one category of ICC and two measurements deteriorated by one category (from good to acceptable or from acceptable to poor). In the rereadings, 12 of 19 measurements had acceptable or good rating of correlation, compared with eight in the initial readings. A statistical trend suggests ICCs were greater for bone measurements (≥ 0.70 in five [initial readings] and eight [rereadings] of nine variables) than for soft-tissue measurements (≥ 0.70 in three [initial readings] and four [rereadings] of 10 variables, $P = .057$).

The categorical variable rereadings show disagreement ($\kappa = -0.34$ to 0.35; Table 5). As with initial readings, categorical variables, such as vaginal shape or presence of levator tear, continued to show poor agreement.

Discussion

Bone pelvimetry measurements were more consistent than soft-tissue mea-

Table 3

Difference between Rereadings from Three Readers

Variable	Mean Distance ($n = 60$) (mm)*	ICC
Bone parameter		
Interspinous distance	108.73 \pm 8.96	0.81
Intertuberous diameter	117.13 \pm 10.21	0.92
Angle of pubic arch with symphysis as apex	85.10 \pm 7.54	0.72
Obstetric conjugate	126.88 \pm 9.43	0.75
Anteroposterior outlet	107.80 \pm 7.90	0.62
Sacral length	122.25 \pm 8.96	0.81
Sacral depth	37.56 \pm 6.93	0.84
Transverse inlet	105.37 \pm 5.35	0.82
Transverse diameter	125.40 \pm 7.12	0.91
Soft-tissue parameter		
Width of levator hiatus	35.15 \pm 6.89	0.76
Width of right levator sling muscle	5.34 \pm 2.84	0.30
Width of left levator sling muscle	6.25 \pm 2.97	0.61
Distance from bladder neck to PCL, measured at rest	21.61 \pm 4.29	0.77
H-line measured at rest	50.05 \pm 7.70	0.43
H-line measured during strain	57.77 \pm 9.27	0.70
Difference between H-line measured rest and during strain	7.72 \pm 8.09	0.78
M-line measured at rest	19.58 \pm 9.32	0.24
M-line measured during strain	34.38 \pm 12.44	0.49
Difference between M-line measured at rest and during strain	14.80 \pm 10.61	0.68

* Data are the mean \pm standard deviation.

Table 4

Summary of Categorical Variables in Initial Trials

Parameter	No/No	No/Yes	Yes/Yes	κ Value*
Is there a gap or tear of the levator sling present?	165 (83.8)	29 (14.7)	3 (1.5)	0.12 (−0.05, 0.28)
Vaginal shape	18 (9.1)	58 (29.3)	122 (61.6)	0.31 (0.16, 0.45)
Is there a gap at the sling insertion to symphysis present?	163 (83.6)	26 (13.3)	6 (3.1)	0.25 (0.04, 0.45)
Is there an internal anal sphincter tear?	186 (93.9)	12 (6.1)	0 (0)	−0.023 (−0.044, −0.003)
Is there an external anal sphincter tear?	177 (90.3)	19 (9.7)	0 (0)	−0.019 (−0.042, 0.005)
Rectocele	142 (72.4)	29 (14.8)	25 (12.8)	0.54 (0.39, 0.69)
Enterocele	193 (99.0)	2 (1.0)	0 (0)	NA
Contour of iliooccygeus	11 (5.7)	17 (8.8)	166 (85.6)	0.52 (0.10, 0.32)

Note.—Potential combined answers to “yes” or “no” questions from two different readers. Unless otherwise noted, values are the number of patients; data in parentheses are the percentages. NA = not applicable.

* Data in parentheses are the 95% confidence intervals.

surements at initial and repeated readings, although this difference was not significant. Some soft-tissue variables, such as resting sagittal measurements of hiatus (M-line and H-

line) and posterior levator plate angles, showed poor correlation despite additional training. Poor delineation of soft-tissue interfaces despite optimized pelvic phased-array imaging

technique likely contributes to the greater interobserver variability of soft-tissue parameter measurements.

Continuous parameters with large values, such as bone pelvimetry measurements, showed the highest overall agreement as a group. Parameters with small values demonstrated high variability. The relative lack of improvement for pelvimetry measurements after additional training is expected, given that these measurements already had high consistency and therefore less room for improvement. Bone parameters tended to have better-defined margins and greater contrast with adjacent soft-tissue structures, particularly fat, enhancing readers' ability to produce reliable measurements. Some variability for bone measurements likely resulted from limited contrast between cortical bone and contiguous hypointense structures (for example, tendons) in areas such as the ischial tuberosities.

Differences between measurements at rest and with strain would be expected to show less variability, since the landmark of each static measurement should be consistent for each reader, including the M-line measurements. This is supported by our data, as the ICC for M-line improved on assessment of the difference between rest and strain.

The literature regarding variability of MR imaging evaluation of pelvic organs is limited (9–13), underscoring the

importance of assessment of measurement reliability in a diagnostic study. There are two components to the variation, one component is the between-image variability and the second component is the reader measurement error (within-image variation). Poor reliability is a result of the disagreement between the two readers (ie, relatively large within-image variation compared with between-image variation).

Even in single-site studies, there may be unacceptable variability in MR imaging measurements. In a study of 10 volunteers, unacceptable variability mainly resulted from high intraobserver variability. There was also high interobserver variability, and the authors recommended strategies to reduce sources of measurement error, such as repeated measurements, repeated examinations, and calibrating observers (12). In our multi-institutional study, the process of training, initial interpretation, retraining, and rereading provided an excellent opportunity to evaluate the variability of pelvic MR measurements among readers with specialized training from different institutions. Our data demonstrate that reproducibility of pelvic MR measurement is improved by targeted training that includes clear agreement about measurement landmarks.

Different measurement software among sites, different MR imagers, inconsistent choice of the same image for measurement among a series of sec-

tions, and variations in the understanding of image landmarks over time could each contribute to the variability. Comparison of overall variability of initial pelvic MR measurements and repeated measurements suggests the existence of technical limitations that extend beyond training of image interpreters. Despite additional training of readers by using techniques to improve interobserver reliability, there was still wide disparity in the effectiveness of the additional training. Several observations suggest underlying reasons for this variation.

In general, the categorical and binary variables showed poor correlation between readers. The inconsistency of these measurements despite additional training may be a result of limitations of the technique (eg, spatial resolution or lack of evacuation of contrast) rather than interpretive errors (inconsistent selection of landmarks or caliper placement). Variations in study acquisition could also lead to differences in interpretations if differences in the performance of the MR imaging affected the image quality or parameters and resulted in improved or worsened landmark visualization of images at one site relative to another despite standardized technique. An interesting finding of our analysis was persistent variability between readers on the rereadings.

Limitations of our study included the lack of inclusion of all potential subjects in the imaging trial; thus, the full spectrum of primiparous women may not be represented. Another potential limitation was selection bias of the subset of MR imaging studies chosen for rereadings. Further, it is theoretically possible, although highly unlikely, that the readers remembered studies from the initial interpretation. Another consideration was that a single day of training may not be adequate for this technique, even for readers with experience in pelvic MR. It is possible that the use of a true T2 imaging sequence may have limited interobserver reliability if boundary artifacts of T2 fast imaging with steady-state precession potentially improved measurements, but this was not evaluated.

Finally, there was some inconsis-

Table 5

Summary of Categorical Variables in Rereadings

Parameter	No/No/No	No/No/Yes	No/Yes/Yes	Yes/Yes/Yes	κ Value*
Is there a gap or tear of the levator sling present?	14 (70.0)	1 (5.0)	5 (25.0)	0 (0)	0.030 (−0.22, 0.28)
Vaginal shape	0 (0)	6 (31.6)	2 (10.5)	11 (57.9)	0.21 (−0.04, 0.46)
Is there a gap at the sling insertion to symphysis present?	0 (0)	13 (76.5)	0 (0)	4 (23.5)	−0.34 (−0.61, −0.07)
Rectocele	11 (55.0)	1 (5.0)	6 (30.0)	2 (10.0)	0.35 (0.10, 0.60)
Enterocoele	20 (100.0)	0 (0)	0 (0)	0 (0)	NA
Contour of iliococcygeus	0 (0)	3 (15.0)	3 (15.0)	14 (70.0)	0.22 (−0.03, 0.47)

Note.—Unless otherwise indicated, values are the number of patients; numbers in parentheses are the percentages. NA = not applicable.

* Numbers in parentheses are the 95% confidence intervals.

tency between readers in the definitions of measurement parameters despite additional training. Three parameters, the distance from bladder neck to PCL with strain, and the angle of levator plate with PCL at rest and with strain, were excluded from statistical analysis mainly owing to inconsistent use of positive and negative signs for the measurements, thus skewing the means of the affected parameters.

In conclusion, our study demonstrated excessive variability of specific pelvic MR measurements performed at separate institutions by different readers. These results have important implications that may limit the use of certain MR measurements for the evaluation and treatment of pelvic floor dysfunction. The evolution of MR imaging techniques with improved distinction of landmarks and greater spatial and contrast resolution, particularly between contiguous soft-tissue structures, will hopefully increase its use in the future.

References

- Borello-France D, Burgio KL, Richter HE, et al. Fecal and urinary incontinence in primiparous women. *Obstet Gynecol* 2006; 108(4):863–872.
- Hodroff MA, Stolpen AH, Denson MA, Bolinger L, Kreder KJ. Dynamic magnetic resonance imaging of the female pelvis: the relationship with the Pelvic Organ Prolapse quantification staging system. *J Urol* 2002; 167(3):1353–1355.
- Singh K, Jakab M, Reid WM, Berger LA, Hoyte L. Three-dimensional magnetic resonance imaging assessment of levator ani morphologic features in different grades of prolapse. *Am J Obstet Gynecol* 2003;188(4): 910–915.
- Fletcher JG, Busse RF, Riederer SJ, et al. Magnetic resonance imaging of anatomic and dynamic defects of the pelvic floor in defecatory disorders. *Am J Gastroenterol* 2003; 98(2):399–411.
- Healy JC, Halligan S, Reznick RH, Watson S, Phillips RK, Armstrong P. Patterns of prolapse in women with symptoms of pelvic floor weakness: assessment with MR imaging. *Radiology* 1997;203(1):77–81.
- Kelvin FM, Maglinte DD, Hale DS, Benson JT. Female pelvic organ prolapse: a comparison of triphasic dynamic MR imaging and triphasic fluoroscopic cystocoloproctography. *AJR Am J Roentgenol* 2000;174(1): 81–88.
- Vanbeckevoort D, Van Hoe L, Oyen R, Ponette E, De Ridder D, Deprest J. Pelvic floor descent in females: comparative study of colpocystodefecography and dynamic fast MR imaging. *J Magn Reson Imaging* 1999; 9(3):373–377.
- Dobben AC, Wiersma TG, Janssen LW, et al. Prospective assessment of interobserver agreement for defecography in fecal incontinence. *AJR Am J Roentgenol* 2005; 185(5):1166–1172.
- Terra MP, Beets-Tan RG, van Der Hulst VP, et al. Anal sphincter defects in patients with fecal incontinence: endoanal versus external phased-array MR imaging. *Radiology* 2005; 236(3):886–895.
- Beets-Tan RG, Morren GL, Beets GL, et al. Measurement of anal sphincter muscles: endoanal US, endoanal MR imaging, or phased-array MR imaging? a study with healthy volunteers. *Radiology* 2001;220(1):81–99.
- Keller TM, Rake A, Michel SC, et al. Obstetric MR pelvimetry: reference values and evaluation of inter- and intraobserver error and intraindividual variability. *Radiology* 2003;227(1):37–43.
- Morren GL, Balasingam AG, Wells JE, Hunter AM, Coates RH, Perry RE. Triphasic MRI of pelvic organ descent: sources of measurement error. *Eur J Radiol* 2005;54(2): 276–283.
- Hetzer FH, Andreisek G, Tsagari C, Sahrbacher U, Weishaupt D. MR defecography in patients with fecal incontinence: imaging findings and their effect on surgical management. *Radiology* 2006;240(2):449–457.
- Morgan DM, Umek W, Stein T, Hsu Y, Guire K, DeLancey JO. Interrater reliability of assessing levator ani muscle defects with magnetic resonance images. *Int Urogynecol J Pelvic Floor Dysfunct* 2007;18(7):773–778.
- Terra MP, Beets-Tan RG, van der Hulst VP, et al. MRI in evaluating atrophy of the external anal sphincter in patients with fecal incontinence. *AJR Am J Roentgenol* 2006; 187(4):991–999.
- Richter HE, Fielding JR, Bradley CS, et al. Endoanal ultrasound findings and fecal incontinence symptoms in women with and without recognized anal sphincter tears. *Obstet Gynecol* 2006;108(6):1394–1401.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20: 37–46.
- Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York, NY: Wiley, 1981; 38–46.

Radiology 2008

This is your reprint order form or pro forma invoice

(Please keep a copy of this document for your records.)

Reprint order forms and purchase orders or prepayments must be received 72 hours after receipt of form either by mail or by fax at 410-820-9765. It is the policy of Cadmus Reprints to issue one invoice per order.

Please print clearly.

Author Name _____
Title of Article _____
Issue of Journal _____ Reprint # _____ Publication Date _____
Number of Pages _____ KB # _____ Symbol Radiology
Color in Article? Yes / No (Please Circle)

Please include the journal name and reprint number or manuscript number on your purchase order or other correspondence.

Order and Shipping Information

Reprint Costs (Please see page 2 of 2 for reprint costs/fees.)

_____ Number of reprints ordered \$ _____
_____ Number of color reprints ordered \$ _____
_____ Number of covers ordered \$ _____
Subtotal \$ _____
Taxes \$ _____

(Add appropriate sales tax for Virginia, Maryland, Pennsylvania, and the District of Columbia or Canadian GST to the reprints if your order is to be shipped to these locations.)

First address included, add \$32 for
each additional shipping address \$ _____

TOTAL \$ _____

Shipping Address (cannot ship to a P.O. Box) Please Print Clearly

Name _____
Institution _____
Street _____
City _____ State _____ Zip _____
Country _____
Quantity _____ Fax _____
Phone: Day _____ Evening _____
E-mail Address _____

Additional Shipping Address* (cannot ship to a P.O. Box)

Name _____
Institution _____
Street _____
City _____ State _____ Zip _____
Country _____
Quantity _____ Fax _____
Phone: Day _____ Evening _____
E-mail Address _____

* Add \$32 for each additional shipping address

Payment and Credit Card Details

Enclosed: Personal Check _____
Credit Card Payment Details _____
Checks must be paid in U.S. dollars and drawn on a U.S. Bank.
Credit Card: VISA Am. Exp. MasterCard
Card Number _____
Expiration Date _____
Signature: _____

Please send your order form and prepayment made payable to:

Cadmus Reprints
P.O. Box 751903
Charlotte, NC 28275-1903

Note: Do not send express packages to this location, PO Box.
FEIN #:541274108

Signature _____ Date _____

Signature is required. By signing this form, the author agrees to accept the responsibility for the payment of reprints and/or all charges described in this document.

Invoice or Credit Card Information

Invoice Address Please Print Clearly

Please complete Invoice address as it appears on credit card statement

Name _____
Institution _____
Department _____
Street _____
City _____ State _____ Zip _____
Country _____
Phone _____ Fax _____
E-mail Address _____

**Cadmus will process credit cards and Cadmus Journal
Services will appear on the credit card statement.**

*If you don't mail your order form, you may fax it to 410-820-9765 with
your credit card information.*

Radiology 2008

Black and White Reprint Prices

Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	\$221	\$233	\$268	\$285	\$303	\$323
5-8	\$355	\$382	\$432	\$466	\$510	\$544
9-12	\$466	\$513	\$595	\$652	\$714	\$775
13-16	\$576	\$640	\$749	\$830	\$912	\$995
17-20	\$694	\$775	\$906	\$1,017	\$1,117	\$1,220
21-24	\$809	\$906	\$1,071	\$1,200	\$1,321	\$1,471
25-28	\$928	\$1,041	\$1,242	\$1,390	\$1,544	\$1,688
29-32	\$1,042	\$1,178	\$1,403	\$1,568	\$1,751	\$1,924
Covers	\$97	\$118	\$215	\$323	\$442	\$555

Color Reprint Prices

Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	\$223	\$239	\$352	\$473	\$597	\$719
5-8	\$349	\$401	\$601	\$849	\$1,099	\$1,349
9-12	\$486	\$517	\$852	\$1,232	\$1,609	\$1,992
13-16	\$615	\$651	\$1,105	\$1,609	\$2,117	\$2,624
17-20	\$759	\$787	\$1,357	\$1,997	\$2,626	\$3,260
21-24	\$897	\$924	\$1,611	\$2,376	\$3,135	\$3,905
25-28	\$1,033	\$1,071	\$1,873	\$2,757	\$3,650	\$4,536
29-32	\$1,175	\$1,208	\$2,122	\$3,138	\$4,162	\$5,180
Covers	\$97	\$118	\$215	\$323	\$442	\$555

International (includes Canada and Mexico)						
# of Pages	50	100	200	300	400	500
1-4	\$272	\$283	\$340	\$397	\$446	\$506
5-8	\$428	\$455	\$576	\$675	\$784	\$884
9-12	\$580	\$626	\$805	\$964	\$1,115	\$1,278
13-16	\$724	\$786	\$1,023	\$1,232	\$1,445	\$1,652
17-20	\$878	\$958	\$1,246	\$1,520	\$1,774	\$2,030
21-24	\$1,022	\$1,119	\$1,474	\$1,795	\$2,108	\$2,426
25-28	\$1,176	\$1,291	\$1,700	\$2,070	\$2,450	\$2,813
29-32	\$1,316	\$1,452	\$1,936	\$2,355	\$2,784	\$3,209
Covers	\$156	\$176	\$335	\$525	\$716	\$905

International (includes Canada and Mexico))						
# of Pages	50	100	200	300	400	500
1-4	\$278	\$290	\$424	\$586	\$741	\$904
5-8	\$429	\$472	\$746	\$1,058	\$1,374	\$1,690
9-12	\$604	\$629	\$1,061	\$1,545	\$2,011	\$2,494
13-16	\$766	\$797	\$1,378	\$2,013	\$2,647	\$3,280
17-20	\$945	\$972	\$1,698	\$2,499	\$3,282	\$4,069
21-24	\$1,110	\$1,139	\$2,015	\$2,970	\$3,921	\$4,873
25-28	\$1,290	\$1,321	\$2,333	\$3,437	\$4,556	\$5,661
29-32	\$1,455	\$1,482	\$2,652	\$3,924	\$5,193	\$6,462
Covers	\$156	\$176	\$335	\$525	\$716	\$905

Minimum order is 50 copies. For orders larger than 500 copies, please consult Cadmus Reprints at 800-407-9190.

Reprint Cover

Cover prices are listed above. The cover will include the publication title, article title, and author name in black.

Shipping

Shipping costs are included in the reprint prices. Domestic orders are shipped via UPS Ground service. Foreign orders are shipped via a proof of delivery air service.

Multiple Shipments

Orders can be shipped to more than one location. Please be aware that it will cost \$32 for each additional location.

Delivery

Your order will be shipped within 2 weeks of the journal print date. Allow extra time for delivery.

Tax Due

Residents of Virginia, Maryland, Pennsylvania, and the District of Columbia are required to add the appropriate sales tax to each reprint order. For orders shipped to Canada, please add 7% Canadian GST unless exemption is claimed.

Ordering

Reprint order forms and purchase order or prepayment is required to process your order. Please reference journal name and reprint number or manuscript number on any correspondence. You may use the reverse side of this form as a proforma invoice. Please return your order form and prepayment to:

Cadmus Reprints
P.O. Box 751903
Charlotte, NC 28275-1903

Note: Do not send express packages to this location, PO Box. FEIN #: 541274108

Please direct all inquiries to:

Rose A. Baynard
800-407-9190 (toll free number)
410-819-3966 (direct number)
410-820-9765 (FAX number)
baynardr@cadmus.com (e-mail)

Reprint Order Forms and purchase order or prepayments must be received 72 hours after receipt of form.