Accuracy and Reliability Assessment of CT and MR Perfusion Analysis Software Using a Digital Phantom[1]

Kohsuke Kudo, MD, PhD
Soren Christensen, PhD
Makoto Sasaki, MD, PhD
Leif Østergaard, MD, PhD
Hiroki Shirato, MD, PhD
Kuniaki Ogasawara, MD, PhD
Max Wintermark, MD, PhD
Steven Warach, MD, PhD
For the Stroke Imaging Repository (STIR) Investigators

Purpose: To design a digital phantom data set for computed tomography (CT) perfusion and perfusion-weighted imaging on the basis of the widely accepted tracer kinetic theory in which the true values of cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT), and tracer arrival delay are known and to evaluate the accuracy and reliability of postprocessing programs using this digital phantom.

Materials and Methods: A phantom data set was created by generating concentration-time curves reflecting true values for CBF (2.5–87.5 mL/100 g per minute), CBV (1.0–5.0 mL/100 g), MTT (3.4–24 seconds), and tracer delays (0–3.0 seconds). These curves were embedded in human brain images. The data were analyzed by using 13 algorithms each for CT and magnetic resonance (MR), including five commercial vendors and five academic programs. Accuracy was assessed by using the Pearson correlation coefficient (r) for true values. Delay-, MTT-, or CBV-dependent errors and correlations between time to maximum of residue function (T_{max}) were also evaluated.

Results: In CT, CBV was generally well reproduced ($r > 0.9$ in 12 algorithms), but not CBF and MTT ($r > 0.9$ in seven and four algorithms, respectively). In MR, good correlation ($r > 0.9$) was observed in one-half of commercial programs, while all academic algorithms showed good correlations for all parameters. Most algorithms had delay-dependent errors, especially for commercial software, as well as CBV dependency for CBF or MTT calculation and MTT dependency for CBV calculation. Correlation was good in T_{max} except for one algorithm.

Conclusion: The digital phantom readily evaluated the accuracy and characteristics of the CT and MR perfusion analysis software. All commercial programs had delay-induced errors and/or insufficient correlations with true values, while academic programs for MR showed good correlations with true values.

Supplemental material: *http://radiology.rsna.org/lookup/suppl/doi:10.1148/radiol.12112618/-/DC1*

[1] From the Division of Ultra-High Field MRI and Department of Radiology (K.K., M.S.), Department of Neurosurgery (K.O.), Iwate Medical University, 19-1 Uchimaru, Morioka 020-8505, Japan; Departments of Neurology and Radiology, University of Melbourne, Royal Melbourne Hospital, Melbourne, Australia (S.C.); Center for Functionally Integrative Neuroscience, Department of Neuroradiology, Århus University Hospital, Århus, Denmark (L.Ø.); Department of Radiology, Hokkaido University Graduate School of Medicine, Sapporo, Japan (H.S.); Department of Radiology, University of Virginia, Charlottesville, Va (M.W.); and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Md (S.W.). Received December 8, 2011; revision requested February 3, 2012; final revision received July 5; accepted July 26; final version accepted September 7. Supported in part by a Grant-in-Aid for Strategic Medical Science Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by the Japan Society for the Promotion of Science through the Funding Program for Next Generation World-Leading Researchers (NEXT Program) initiated by the Council for Science and Technology Policy. **Address correspondence to** K.K. (e-mail: *kokudo@iwate-med.ac.jp*).

Computed tomography (CT) perfusion is widely used for acute ischemic stroke (1–3), hemodynamic ischemia (4), subarachnoid hemorrhage (5), and brain tumors (6). Dynamic susceptibility contrast perfusion-weighted imaging also has been used in studies on acute ischemic stroke, such as the identification of ischemic penumbra (7,8) and the prediction of final infarction (9,10). Recently, a variety of postprocessing programs and algorithms for CT perfusion and perfusion-weighted imaging have been made available by CT and magnetic resonance (MR) imaging unit manufacturers, third-party workstation vendors, and academic groups. However, there are substantial differences between these programs and algorithms in terms of their maps and quantitative values in CT perfusion and perfusion-weighted imaging (11–14). It is difficult to determine the accuracy and reliability of each program, because precise implementations of analysis algorithms are generally not open to users, particularly in the case of commercial programs. Information about the type of algorithm, such as singular value decomposition (SVD) or maximum slope algorithm, is usually provided, but this information has little bearing on the correctness of its implementation and its accuracy. In addition, the accuracy and reliability of these programs have not been subject to standardized quality control. The importance of standardizing perfusion processing has recently been identified as a key factor in advancing stroke imaging (15).

The purpose of the present study was twofold: *(a)* to design a digital phantom data set for CT perfusion and perfusion-weighted imaging on the basis of the widely accepted tracer kinetic theory in which the true values of cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT), and tracer arrival delay are known and *(b)* to evaluate the accuracy and reliability of postprocessing programs by using this digital phantom.

Advances in Knowledge

- Accuracy and reliability of perfusion analysis software, including tracer-delay sensitivity, can be objectively evaluated by using a digital phantom data set that is based on the widely accepted tracer kinetic theory.
- Correlation with true values embedded in digital phantom data and delay sensitivity were different among software packages and algorithms; however, the academic software had better correlations ($r > 0.9$) with true values, especially for MR.
- Time to maximum of residue function (T_{max}) values in MR were similar in most software packages, and correlations with each other were high ($r > 0.9$) although none of the commercial software studied here could be used to produce a T_{max} map.

Materials and Methods

Curve Generation for Digital Phantom

Concentration-time curves for the artery (arterial input function), vein (venous output function), and brain tissue were generated by using previously described methods (Appendix E1 [online]) (16,17).

Seven MTT values (3.4, 4.0, 4.8, 6.0, 8.0, 12.0, and 24.0 seconds) and five CBV values (1.0, 2.0, 3.0, 4.0, and 5.0 mL/100 g) were used for this curve simulation. The CBF values were determined from each pair of CBV and MTT values by using the central volume principle (CBF = CBV/MTT) (18). For example, CBF values were 2.5, 5.0, 7.5, 10.0, 12.5, 15.0, and 17.5 mL/100 g per minute when CBV was 1.0 mL/100 g, and CBF values were 12.5, 25.0, 37.5, 50.0, 62.5, 75.0, and 87.5 mL/100 g per minute when CBV was 5.0 mL/100 g. Seven increments of tracer delay (0, 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 seconds) were achieved by shifting the concentration-time curve of tissue, or $C(t)$, relative to the arterial input function. In total, 735 combinations of tissue curves were generated from three types of residue function ($R[t]$), seven values of MTT, five values of CBV, and seven delays.

Implication for Patient Care

- Physicians should be careful about the accuracy of analysis software they use, because most of the algorithms in commercial software have delay-induced errors and have additional sources of error that are yet to be elucidated.

Data Structure

The digital phantom data set contained 16 section locations both for CT perfusion and perfusion-weighted imaging (Fig 1). Section 1 contained the arterial input function and venous output function, and tissue curves were embedded in sections 2–16. Matrixes of these sections were 512 × 512 for CT perfusion and 256 × 256 for perfusion-weighted imaging. Each tissue section contained 7 × 7 quadratic tiles. Each tile was encoded with increasing MTT (decreasing CBF) across columns (left to right) and increasing delays across rows. The data were exported to Digital Imaging and

Published online before print
10.1148/radiol.12112618 **Content code:** NR

Radiology 2013; 267:201–211

Abbreviations:
ANOVA = analysis of variance
CBF = cerebral blood flow
CBV = cerebral blood volume
MTT = mean transit time
PMA = Perfusion Mismatch Analyzer
$R(t)$ = residue function
SVD = singular value decomposition
T_{max} = time to maximum of $R(t)$

Author contributions:
Guarantor of integrity of entire study, K.K.; 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, K.K., L.Ø.; experimental studies, K.K., S.C., M.S.; statistical analysis, K.K.; and manuscript editing, all authors

Funding:
S.W. is an employee of the National Institute of Neurological Disorders and Stroke, National Institutes of Health.

Conflicts of interest are listed at the end of this article.

Figure 1

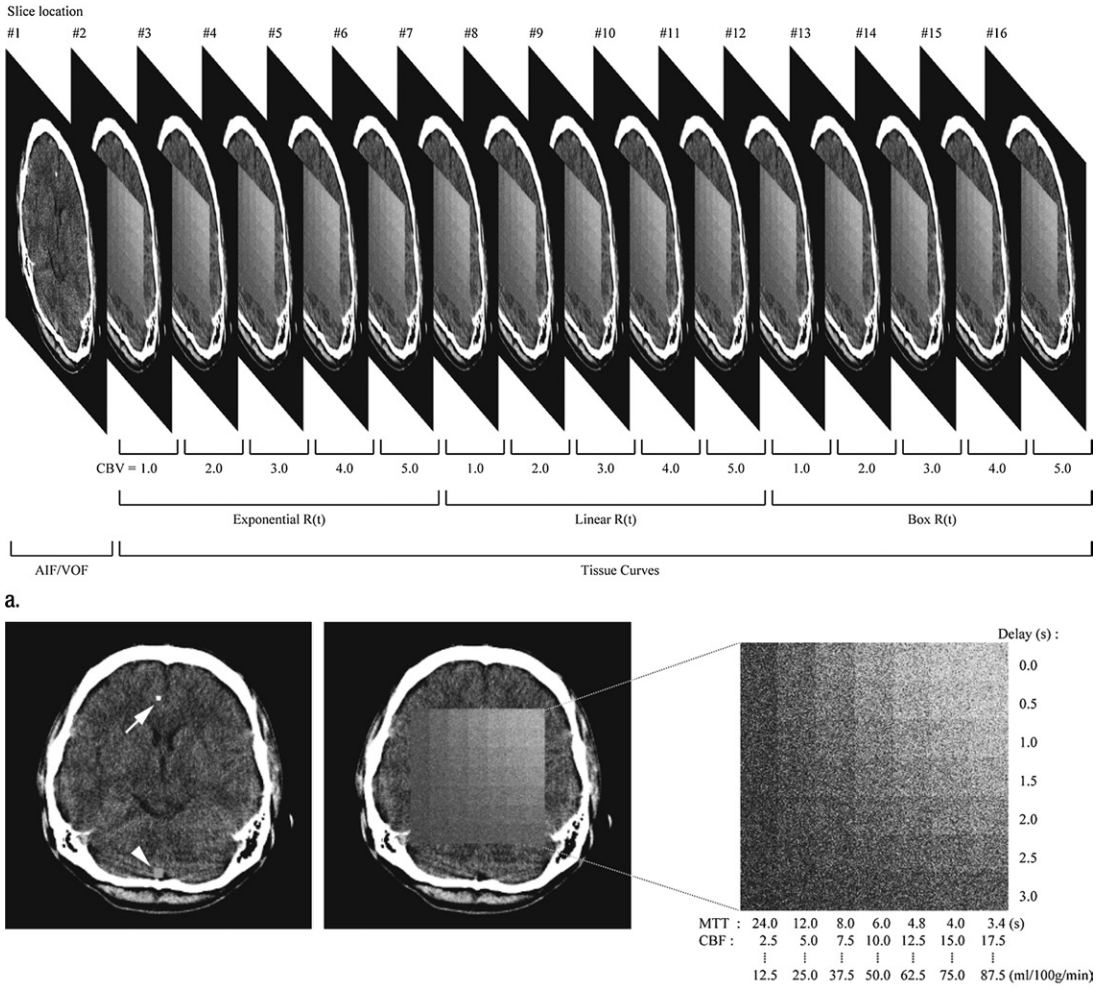

Figure 1: Data structure of digital phantom for CT. Digital phantom data consist of 16 sections (**a**). Image shows section *(Slice)* location. Section 1 (**b**, left) contains curves for arterial input function (arrow on **b**, left) and venous output function (arrowhead on **b**, left). Tissue curves are embedded in sections 2–16 (**b**, middle). The tissue section contains 7 × 7 quadratic tiles (**b**, right), which has a different delay and MTT. The first five sections (sections 2–6) have curves of exponential $R(t)$, with each section having a particular value of CBV (1, 2, 3, 4, and 5 mL/100 g, respectively) (**a**). The next five sections (sections 7–11) and the final five sections (sections 12–16) have curves of linear $R(t)$ and box-shaped $R(t)$, respectively.

Communications in Medicine format, enabling import into all software packages used in the present study.

Data Analysis

The phantom data were postprocessed by using five commercially available software packages (GE Healthcare, Waukesha, Wis; Hitachi Medical Systems, Chiba, Japan; Philips Medical Systems, Best, the Netherlands; Siemens Healthcare, Erlangen, Germany; and Toshiba Medical Systems, Tokyo, Japan) and two academic programs (Perfusion Mismatch Analyzer [PMA], Acute Stroke Imaging Standardization Group [ASIST], Japan, *http://asist.umin.jp/index-e.htm* [11,12]; and Stroketool, Digital Image Solutions, Germany, *http://www.digitalimagesolutions.de/*) for CT perfusion (Table 1). Also used were four commercially available software packages (GE Healthcare, Hitachi Medical Systems, Philips Medical Systems, and Siemens Healthcare) and five academic programs, including Echoplanar Imaging Thrombolysis Evaluation Trial, or EPITHET (Melbourne Health, Melbourne, Australia) (19); perfusion graphical user interface, or Penguin (Danish National Research Foundation, Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark; *http://www.cfin.au.dk/software/penguin*); RAPID (Stanford University, Stanford, Calif)

Table 1

Software for CT Perfusion

Abbreviation	Manufacturer or Access	Software Name and Version	Algorithm
G1	GE Healthcare	CT Perfusion 3	Deconvolution, SVD
G2	GE Healthcare	CT Perfusion 4	Deconvolution, SVD
H	Hitachi Medical Systems	Perfusion Analysis, version 3.0	Deconvolution, inverse filter
P1	Philips Medical Systems	Extended Brilliance Workspace, version 4.0	Deconvolution, standard SVD
P2	Philips Medical Systems	Extended Brilliance Workspace, version 4.0	Deconvolution, block-circulant SVD
S1	Siemens Healthcare	Syngo MMWP VE36A	Maximum slope
S2	Siemens Healthcare	Syngo MMWP VE36A	Deconvolution, least mean square
T1	Toshiba Medical Systems	CBP Study Ph8	Deconvolution, box-modulated transfer function
T2	Toshiba Medical Systems	CBP Study Ph8	Deconvolution, reformulated SVD
PM1	Acute Stroke Imaging Standardization Group	PMA, version 3.0.0.0	Deconvolution, standard SVD
PM2	Acute Stroke Imaging Standardization Group	PMA, version 3.0.0.0	Deconvolution, block-circulant SVD
ST1	Digital Image Solutions	Stroketool, 2010 version	Deconvolution, standard SVD
ST2	Digital Image Solutions	Stroketool, 2010 version	Deconvolution, optimized SVD

Table 2

Software for Perfusion-weighted Imaging

Abbreviation	Manufacturer or Access	Software Name and Version	Algorithm
G	GE Healthcare	AW 4.4, BrainStat	Gamma Fitting
H	Hitachi Medical Systems	V2.0A SP4	Gamma Fitting
P	Philips Medical Systems	R.2.6.1	Deconvolution, block-circulant SVD
S	Siemens Healthcare	VB17	Deconvolution, standard SVD
EP	Melbourne University	Echoplanar Imaging Thrombolysis Evaluation Trial*	Deconvolution, standard SVD
PG1	Danish National Research Foundation	Perfusion graphical user interface*	Deconvolution, standard SVD
PG2	Danish National Research Foundation	Perfusion graphical user interface*	Deconvolution, oscillation-index optimized SVD
PM1	Acute Stroke Imaging Standardization Group	PMA ,version 3.0.0.0	Deconvolution, standard SVD
PM2	Acute Stroke Imaging Standardization Group	PMA, version 3.0.0.0	Deconvolution, block-circulant SVD
RP1	Stanford University	RAPID, V2.1	Deconvolution, standard SVD
RP2	Stanford University	RAPID, V2.1	Deconvolution, circular SVD
ST1	Digital Image Solutions	Stroketool, V2.3	Deconvolution, standard SVD
ST2	Digital Image Solutions	Stroketool, V2.3	Deconvolution, optimized SVD

* Information about the version was not available.

(20); PMA; and Stroketool (Digital Image Solutions) for perfusion-weighted imaging (Table 2). CBF, CBV, and MTT maps were generated by each package, except for the Echoplanar Imaging Thrombolysis Evaluation Trial program, which was limited to time to maximum of $R(t)$ (T_{max}) maps. All the academic programs for perfusion-weighted imaging also included generation of T_{max} maps.

Region-of-interest measurements were automatically performed by using in-house software. The average pixel value for each of the 735 tiles was measured by using a square region of interest of 28 × 28 pixels in size for CT perfusion and one of 12 × 12 pixels in size for perfusion-weighted imaging (excluding the edge pixels, the pixel value of which might be affected by the surrounding pixels during smoothing procedures). To evaluate the accuracy of the software packages, the Pearson correlation coefficient r was calculated to determine correlation between each program's results and the true values of CBF, CBV, and MTT. A good correlation was defined as one with $r > 0.9$. The slope and intercept of linear regression were also determined to evaluate the relationship between the values measured from each program and the true values. Differences among algorithms were tested by using one-way repeated-measures analysis of variance (ANOVA), and 95% confidence intervals for the slope and intercept were calculated. A two-way repeated-measures

Figure 2

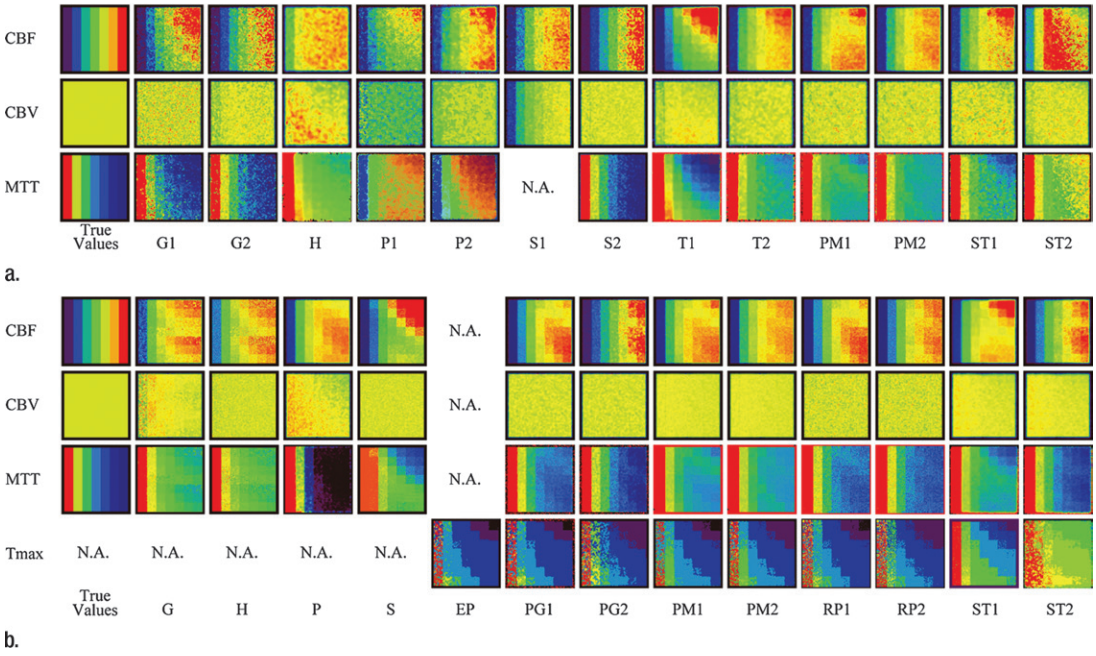

Figure 2: Perfusion maps. True-value and perfusion maps of digital phantom (section 16, box-shaped $R(t)$, and CBV [5 mL/100 g]) are shown with identical color lookup table (not shown). The color scale is automatically adjusted for visual assessment in each map of CBF, CBV, and MTT. A constant window level and window width (8 and 16 seconds, respectively) are used for T_{max}. In the true-value maps of CBF and MTT, seven columns with different CBF or MTT show different colors. However, there is no vertical color gradient, because all seven rows have identical CBF or MTT and only the delay is different. The true-CBV image is uniform in color, because all 49 tiles have the same CBV value in a section. **(a)** CT perfusion. Delay-dependent vertical gradation is apparent in most of CBF (G1, P1, P2, T1, T2, PM1, ST1, and ST2) and MTT (G1, H, P1, P2, T1, T2, PM1, PM2, ST1, and ST2). In CBV, a delay-dependent vertical gradient is observed in H and T1. MTT dependency is also noted in G2, H, S1, T1, T2, and ST2 as a horizontal gradient. **(b)** Perfusion-weighted imaging. Delay dependency is apparent in most algorithms for CBF and MTT maps. Delay-dependent vertical gradient on CBV map is observed for P. MTT dependency is also noted for G, S1, and ST2 as a horizontal gradient. *N.A.* = not available.

ANOVA was used to detect the bias of parameter estimates across the different true delay, CBV, and MTT values. To examine the delay-induced error in CBF, CBV, and MTT calculations for each program, delay was modeled as a main factor in the ANOVA. To examine the CBV dependency of CBF and MTT calculation, true CBV was modeled as a main factor in their calculation. Then, true MTT was used as a main factor for CBV calculation. Dependency on different $R(t)$ values was also evaluated by using one-way ANOVA.

The Pearson correlation coefficient r was calculated for T_{max} values for perfusion-weighted imaging between any pair of algorithms. A good correlation was defined as one with $r > 0.9$. The absolute values of T_{max} were compared among the algorithms by using one-way repeated-measures ANOVA followed by a pairwise comparison with the Holm-Sidak test. The Pearson correlation coefficient r between T_{max} and simulated delay or MTT was also calculated. The slope and intercept of linear regression were determined to evaluate the relationship between the T_{max} values from each program and simulated delay or MTT.

A difference with a P value of less than .01 was treated as significant for all tests.

Results

In CT perfusion, CBF maps (Fig 2a) showed delay-dependent variation for G1, P1, P2, T1, T2, PM1, ST1, and ST2 as a color gradient in the vertical direction; see Table 1 for an explanation of the abbreviations used here, in Figure 2a, and hereafter. In contrast, the color gradient was not clear in maps for G2, H, S1, S2, and PM2, which indicates delay insensitivity of these programs. Similarly, a delay-dependent vertical gradient was noted in MTT maps of G1, H, P1, P2, T1, T2, PM1, PM2, ST1, and ST2, whereas no vertical difference was observed in maps for G2 and S2. Delay-dependent vertical gradients were observed in CBV maps for H and T1. MTT dependencies were also noted in maps for G2, H, S1, T1, T2, and ST2 as a horizontal gradient.

In perfusion-weighted imaging, CBF maps (Fig 2b) showed delay-dependent variation for G, H, P, S, PG1, PG2, PM1, RP1, and ST1, while maps for PM2, RP2, and ST2 did not show delay sensitivity; see Table 2 for an

explanation of the abbreviations used here, in Figure 2b, and hereafter. Similarly, delay-dependency was clearly seen in MTT maps for S and ST1 and was observed to some extent in maps for G, H, PG1, PG2, PM1, PM2, RP1, and ST2. Delay-dependent vertical gradients in CBV maps for G and P were observed. MTT dependencies were also noted in maps for G, P, ST1, and ST2 as a horizontal gradient.

Correlation with True Values

In CT perfusion, a good correlation ($r > 0.9$) was achieved for CBF in seven of 13 algorithms (G1, P2, S1, T2, PM1, PM2, and ST1) (Table 3). Most programs showed a good correlation with true CBV, except for one algorithm (S1). Only four (H, T1, T2, and PM2) showed good correlation with true MTT. Only two (T2 and PM2) of 13 algorithms showed good correlation for all three parameters (ie, CBF, CBV, and MTT).

In perfusion-weighted imaging, all the academic programs showed a good correlation ($r > 0.9$) for CBF, CBV, and MTT (Table 4). In contrast, only two, three, and one of the four commercial programs showed a good correlation for CBF, CBV, and MTT, respectively.

Ideally, the slope and intercept of linear regression with true values are expected to be 1.0 and zero, respectively. However, 95% confidence intervals of slope and intercept were away from 1.0 and zero in most of the algorithms, indicating that the absolute values calculated by using the programs were subject to scaling and offsetting errors (Tables 3, 4). In addition, these slopes and intervals significantly varied among the software packages according to ANOVA for all three parameters (CBF, CBV, and MTT) of both CT and MR.

Analysis of Error Sources

In CT perfusion, only four of 13 algorithms (H, S1, S2, and PM2) did not show significant delay dependency on CBF (Table 5). Most algorithms produced a CBF that depended on true CBV, except for two algorithms (H and

Table 3

Correlation and Linear Regression with True Value for CT Perfusion

Software	Correlation Coefficient, *r*			Slope			Intercept		
	CBF	CBV	MTT	CBF	CBV	MTT	CBF	CBV	MTT
G1	0.930*	0.987*	0.882	0.858 (0.761, 0.956)	0.858 (0.847, 0.882)	0.560 (0.471, 0.650)	5.694 (4.512, 6.876)	0.085 (0.059, 0.117)	2.248 (1.260, 3.237)
G2	0.884	0.979*	0.867	0.601 (0.475, 0.727)	0.887 (0.877, 0.956)	0.578 (0.470, 0.685)	4.623 (3.252, 5.993)	0.111 (0.052, 0.133)	4.510 (3.254, 5.765)
H	0.841	0.964*	0.921*	1.033 (0.929, 1.137)†	2.290 (2.211, 2.414)	0.171 (0.153, 0.190)	63.678 (44.275, 83.081)	−0.046 (−0.072, 0.075)‡	2.555 (2.310, 2.800)
P1	0.881	0.920*	0.849	0.982 (0.896, 1.069)†	2.266 (2.189, 2.525)	0.567 (0.465, 0.670)	18.619 (12.784, 24.454)	−0.080 (−0.288, 0.109)‡	5.115 (4.188, 6.043)
P2	0.903*	0.966*	0.896	0.383 (0.364, 0.403)	0.849 (0.834, 0.879)	0.626 (0.556, 0.696)	7.971 (5.755, 10.188)	0.722 (0.700, 0.818)	6.634 (5.437, 7.831)
S1	0.938*	0.790	NA	1.227 (1.115, 1.338)	1.660 (1.606, 2.099)	NA	10.994 (6.164, 15.825)	−0.533 (−0.590, −0.297)	NA
S2	0.869	0.949*	0.870	1.745 (1.376, 2.113)	2.197 (2.131, 2.423)	0.651 (0.503, 0.799)	4.164 (1.675, 6.653)	0.001 (−0.028, 0.066)‡	4.106 (2.418, 5.794)
T1	0.863	0.957*	0.922*	0.648 (0.569, 0.727)	1.018 (0.994, 1.113)†	0.677 (0.650, 0.704)	2.524 (1.381, 3.668)	−0.397 (−0.411, −0.359)	3.930 (3.167, 4.693)
T2	0.921*	0.974*	0.947*	0.491 (0.455, 0.528)	1.018 (0.997, 1.079)†	0.537 (0.496, 0.578)	7.629 (4.858, 10.401)	−0.331 (−0.346, −0.277)	4.091 (3.376, 4.807)
PM1	0.913*	0.983*	0.886	0.796 (0.734, 0.858)	0.914 (0.894, 0.953)	0.550 (0.433, 0.666)	14.256 (9.907, 18.604)	−0.240 (−0.247, −0.207)	4.850 (3.948, 5.751)
PM2	0.923*	0.983*	0.907*	0.849 (0.778, 0.920)	0.914 (0.894, 0.953)	0.575 (0.466, 0.684)	13.378 (9.378, 17.379)	−0.240 (−0.247, −0.207)	4.736 (3.826, 5.646)
ST1	0.917*	0.960*	0.802	0.607 (0.568, 0.646)	1.223 (1.178, 1.319)	0.557 (0.421, 0.693)	9.600 (6.285, 12.916)	−0.529 (−0.548, −0.503)	7.596 (6.236, 8.956)
ST2	0.844	0.940*	0.673	0.318 (0.244, 0.393)	1.107 (1.060, 1.214)	0.291 (0.192, 0.389)	14.236 (8.957, 19.514)	−0.229 (−0.240, −0.171)	11.753 (10.432, 13.073)

Note.—The software abbreviations are defined in Tables 1 and 2. Numbers in parentheses for slope and intercept are 95% confidence intervals. $y = ax + b$, x = true value, y = measured value, a = slope, b = intercept. NA = not available.

* Values represent good correlations ($r > 0.9$).

† The interval of the slope includes 1.0.

‡ The interval of the intercept includes zero.

Table 4

Correlation and Linear Regression with True Value for Perfusion-weighted Imaging

Software	Correlation Coefficient, *r*			Slope			Intercept		
	CBF	CBV	MTT	CBF	CBV	MTT	CBF	CBV	MTT
G	0.810	0.569	0.849	1.719 (1.206, 2.232)	0.615 (0.419, 0.595)	0.213 (0.146, 0.279)	36.488 (33.220, 39.756)	5.151 (4.958, 5.929)	4.000 (3.491, 4.509)
H	0.905*	0.984*	0.881	0.479 (0.453, 0.506)	11.686 (11.425, 12.019)	1.001 (0.812, 1.189)†	13.772 (9.567, 17.978)	−0.793 (−1.065, 0.242)‡	18.398 (16.039, 20.758)
P	0.917*	0.992*	0.932*	4.460 (4.206, 4.713)	6.153 (5.930, 6.203)	1.900 (1.827, 1.974)	88.511 (62.052, 114.970)	0.494 (0.447, 0.640)	−2.636 (−5.390, 0.118)‡
S	0.890	0.996*	0.716	2.014 (1.801, 2.227)	3.651 (3.610, 3.723)	0.951 (0.834, 1.068)†	16.248 (12.909, 19.586)	0.065 (0.047, 0.095)	21.600 (19.048, 24.153)
EP	NA	NA	NA	NA	NA	NA	NA	NA	NA
PG1	0.928*	1.000*	0.970*	0.330 (0.298, 0.362)	3.697 (3.687, 3.717)	0.861 (0.828, 0.895)	3.731 (2.666, 4.796)	−0.048 (−0.077, −0.028)	2.602 (1.983, 3.220)
PG2	0.941*	1.000*	0.977*	0.439 (0.378, 0.500)	3.697 (3.687, 3.717)	0.896 (0.867, 0.925)	2.141 (1.576, 2.706)	−0.048 (−0.077, −0.028)	1.670 (1.125, 2.214)
PM1	0.919*	0.997*	0.954*	1.652 (1.536, 1.769)	1.828 (1.808, 1.835)	0.701 (0.625, 0.777)	26.378 (18.057, 34.698)	−0.414 (−0.442, −0.278)	3.743 (3.016, 4.470)
PM2	0.926*	0.997*	0.957*	1.683 (1.564, 1.802)	1.828 (1.808, 1.835)	0.719 (0.643, 0.794)	25.240 (17.233, 33.247)	−0.414 (−0.442, −0.278)	3.804 (3.074, 4.534)
RP1	0.927*	0.999*	0.955*	1.253 (1.131, 1.374)	2.195 (2.189, 2.206)	0.724 (0.665, 0.784)	15.369 (11.306, 19.431)	0.020 (0.002, 0.038)	2.892 (2.236, 3.549)
RP2	0.935*	0.999*	0.953*	1.341 (1.199, 1.482)	2.195 (2.189, 2.206)	0.706 (0.639, 0.774)	14.517 (10.997, 18.037)	0.020 (0.002, 0.038)	2.759 (2.100, 3.418)
ST1	0.913*	0.996*	0.960*	1.228 (1.153, 1.303)	1.728 (1.691, 1.746)	0.814 (0.774, 0.853)	21.040 (14.126, 27.955)	−0.104 (−0.113, −0.072)	3.431 (2.652, 4.210)
ST2	0.951*	0.996*	0.977*	1.874 (1.678, 2.071)	1.728 (1.691, 1.746)	0.838 (0.802, 0.875)	13.676 (9.908, 17.443)	−0.104 (−0.113, −0.072)	1.979 (1.299, 2.660)

Note.—The software abbreviations are defined in Tables 1 and 2. Numbers in parentheses for slope and intercept are 95% confidence intervals. $y = ax + b$, x = true value, y = measured value, a = slope, b = intercept. NA = not available.

* Values represent good correlations ($r > 0.9$).

† The interval of the slope includes 1.0.

‡ The interval of the intercept includes zero.

P1). However, because of the interaction between delay and true CBV, program H was virtually dependent on both delay and true CBV, where delay dependency was observed with low CBV (1.0 mL/100 g), and CBV dependency was also noted with small amounts of delay (0, 1.0, and 1.5 seconds). The CBV of five algorithms showed delay dependency, whereas the remaining eight algorithms (G1, P1, P2, S1, S2, T2, PM1, and PM2) showed no delay dependency. On the other hand, all the algorithms showed significant CBV differences across the true MTT values, indicating that all the algorithms were susceptible to MTT differences on the basis of erroneous CBV calculation. For MTT, 11 were biased by delay and only two algorithms (S2 and PM2) were delay insensitive. However, both were dependent on CBV. Six algorithms (G2, H, P2, PM1, ST1, and ST2) also had true CBV dependency.

In perfusion-weighted imaging, 10 of 13 algorithms showed significant delay dependency for CBF (Table 6). Only two algorithms (PM2 and RP2) did not show any significant delay dependency. Most algorithms produced a CBF that depended on true CBV, except for one algorithm (PG1). CBV calculations in four algorithms showed delay dependency, whereas seven algorithms (H, PG1, PG2, PM1, PM2, RP1, and RP2) did not show any delay dependency. On the other hand, all the algorithms showed significant CBV differences across true MTT values, indicating that all the algorithms were susceptible to MTT differences on the basis of erroneous CBV calculation. For MTT, 10 algorithms were biased by delay and only two algorithms (PM2 and RP2) were delay insensitive.

Significant dependencies on different types of $R(t)$ were noted in almost all algorithms for both CT perfusion and perfusion-weighted imaging.

Correlation of T_{max} for Perfusion-weighted Imaging

Almost all the pairs, except for one algorithm (ST1), showed good correlation for T_{max} ($r > 0.9$) (Table 7). Absolute values of T_{max} showed a significant

difference with ANOVA, and a post hoc test showed that two algorithms (ST1 and ST2) showed significant differences from others. Correlations with delay or MTT and slopes of linear regression for delay or MTT were similar among all the algorithms, except for one algorithm (ST1). Intercepts for delay were also similar among most algorithms, except for two algorithms (ST1 and ST2); however, intercepts for MTT were more variable than delay among the algorithms.

Discussion

In the current study, we demonstrated that the accuracy and reliability of an analysis program can be objectively assessed by using a digital phantom data set obtained on the basis of the widely accepted tracer kinetic theory in which true values of CBF, CBV, MTT, and tracer arrival delay are known. Delay-induced errors in CBF, CBV, and MTT are easily recognized visually. Statistical evaluation with region of interest measurements is also possible with the digital phantom, and the accuracy of the perfusion analysis can be determined by calculating correlations with true values, as well as error sources such as delay sensitivity and/or MTT or CBV dependencies.

In CT perfusion, only two algorithms (T2 and PM2) showed accurate results; however, algorithm T2 showed delay dependencies on CBF and MTT, and these can lead to potential errors in CBF and MTT calculations. In contrast, any delay dependencies were not shown in PM2, in which the circular type of SVD, developed as a delay-insensitive algorithm in perfusion-weighted imaging (17), was used. In perfusion-weighted imaging, all the academic programs and only one commercial program showed good correlations ($r > 0.9$) for all three parameters. However, most of these algorithms showed delay dependencies on CBF and MTT. Only two algorithms (PM2 and RP2) did not show any delay dependencies for CBF, CBV, and MTT. For these two algorithms, a circular type of SVD, as described above, is used. For the other programs, use of the same types of circular SVD (P, PG2, and ST2) was also stated; however, our phantom analysis revealed that despite this statement, these programs actually had delay-sensitive errors. When we compared CT

Table 5

***P* Values of Repeated-measures ANOVA for CT Perfusion**

	Delay Dependency			True CBV or True MTT Dependency			*R*(*t*) Dependency		
Software	CBF	CBV	MTT	CBF	CBV	MTT	CBF	CBV	MTT
G1	< .001	.288*	<.001	<.001	<.001	.113*	<.001	<.001	<.001
G2	< .001	< .001	<.001	.002	<.001	.006	<.001	<.001	<.001
H	.136†	< .001‡	<.001	.671†	<.001‡	<.001	<.001	<.001	<.001
P1	< .001	.863*	<.001	.964*	<.001	.736*	<.001	<.001	.603*
P2	< .001‡	.033*	<.001‡	<.001‡	<.001	<.001‡	<.001	<.001	<.001
S1	.643*	.474†	NA	<.001	<.001‡	NA	<.001	<.001	NA
S2	.594*	.469*	.471*	<.001	<.001	.001	<.001	<.001	<.001
T1	< .001‡	<.001‡	<.001	<.001‡	<.001‡	.033*	<.001	<.001	<.001
T2	< .001	.357*	<.001	<.001	<.001	.287*	<.001	<.001	<.001
PM1	< .001	.509*	<.001	.001	<.001	<.001	<.001	<.001	<.001
PM2	.631*	.509*	.318*	.005	<.001	.001	<.001	<.001	<.001
ST1	< .001‡	<.001‡	<.001‡	<.001‡	<.001‡	<.001‡	<.001	<.001	<.001
ST2	< .001	<.001‡	<.001	<.001	<.001‡	<.001	<.001	<.001	.018*

Note.—The software abbreviations are defined in Tables 1 and 2. For CBF and MTT, true CBV dependency is presented. For CBV, true MTT dependency is shown. NA = not available.

* No significant dependency.

† No significant dependency but significant interaction between delay and true CBV or true MTT.

‡ Significant interaction between delay and true CBV or true MTT.

Table 6

***P* Values of Repeated-measures ANOVA for Perfusion-weighted Imaging**

	Delay Dependency			True CBV or True MTT Dependency			*R*(*t*) Dependency		
Software	CBF	CBV	MTT	CBF	CBV	MTT	CBF	CBV	MTT
G	<.001*	<.001*	<.001	<.001*	<.001*	.037†	<.001	<.001	.015†
H	<.001*	.202†	<.001*	<.001*	<.001	.996‡	<.001	<.001	<.001
P	<.001*	<.001*	<.001	<.001*	<.001*	<.001	<.001	<.001	<.001
S	<.001*	<.001	<.001*	<.001*	<.001	<.001*	<.001	<.001	<.001
EP	NA	NA	NA	NA	NA	NA	NA	NA	NA
PG1	<.001	.533†	<.001*	.028†	<.001	.001*	<.001	<.001	<.001
PG2	<.001*	.533†	<.001	<.001*	<.001	.058†	<.001	<.001	<.001
PM1	<.001	.056†	<.001*	<.001	<.001	<.001*	<.001	<.001	<.001
PM2	.557†	.056†	.544†	<.001	<.001	<.001	<.001	<.001	<.001
RP1	<.001*	.485†	<.001*	<.001*	<.001	.053‡	<.001	.087†	<.001
RP2	.035†	.485†	.317†	<.001	<.001	.017†	<.001	.087†	<.001
ST1	<.001*	<.001*	<.001*	<.001*	.005*	<.001*	<.001	<.001	<.001
ST2	<.001*	<.001*	<.001*	<.001*	.005*	.018‡	<.001	<.001	<.001

Note.—The software abbreviations are defined in Tables 1 and 2. For CBF and MTT, true CBV dependency is presented. For CBV, true MTT dependency is shown. NA = not available.

* Significant interaction between delay and true CBV or true MTT.

† No significant dependency.

‡ No significant dependency but significant interaction between delay and true CBV and true MTT.

Table 7

Correlation and Linear Regression of T_{max} Compared with Other Algorithms, Delay, or MTT

	Correlation Coefficient, *r*											Slope		Intercept	
Software Algorithm	EP	PG1	PG2	PM1	PM2	RP1	RP2	ST1	ST2	Delay	MTT	Delay	MTT	Delay	MTT
EP	. . .	0.979*	0.941*	0.943*	0.961*	0.967*	0.937*	0.492†	0.929‡	0.376	0.654	0.911	0.235	3.597	2.873
PG1	. . .	. . .	0.956*	0.949*	0.968*	0.955*	0.924*	0.487†	0.939‡	0.383	0.651	0.922	0.233	3.566	2.880
PG2	. . .	. . .	. . .	0.928‡	0.960‡	0.931*	0.932*	0.491†	0.984‡	0.388	0.648	1.011	0.250	3.174	2.465
PM1	. . .	. . .	. . .	. . .	0.970*	0.941*	0.925*	0.547†	0.918‡	0.435	0.683	1.001	0.233	3.768	3.195
PM2	. . .	. . .	. . .	. . .	. . .	0.957*	0.950*	0.539†	0.949‡	0.408	0.691	1.020	0.256	3.747	2.999
RP1	. . .	. . .	. . .	. . .	. . .	. . .	0.985*	0.598†	0.915‡	0.358	0.733	0.918	0.279	3.733	2.627
RP2	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.616†	0.918‡	0.381	0.733	1.029	0.294	3.549	2.480
ST1	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.482†	0.121	0.907	1.126	1.251	8.757	−0.674
ST2	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.420	0.629	0.982	0.218	8.812	8.343

Note.—The software algorithm abbreviations are defined in Tables 1 and 2.

* Value represents good correlations ($r > 0.9$).

† Significant difference, determined with post hoc pairwise comparisons (Holm-Sidak test).

‡ Significant difference, determined with post hoc pairwise comparisons (Holm-Sidak test), but the value represents good correlations ($r > 0.9$).

perfusion and perfusion-weighted imaging, correlations were generally better in perfusion-weighted imaging than in CT perfusion. The differences of phantom configuration were signal-to-noise ratio, temporal resolution, and scan duration. These factors are intrinsic for perfusion methods; for example, low signal-to-noise ratio, longer temporal resolution, and short scan duration for CT perfusion were generally a result from the aim of reduction for radiation dose. Although CT perfusion has linearity between tracer concentration and signal, accuracy is affected by these factors, compared with perfusion-weighted imaging.

According to a previous CT perfusion study with the use of G1, H, P1, S1, and T1 algorithms, overestimation of the infarcted area because of delay-sensitive CBF reduction was found for G1, P1, and T1 (12). In fact, these three algorithms showed delay sensitivity in our digital phantom analysis. Similarly, it has been reported that G1, H, P1, and T1 show a delay-sensitive MTT increase, and our results with phantoms suggest that these algorithms actually have delay-sensitive errors. Most CT unit manufacturers have recently released new versions of their analysis algorithms (G2, P2, S2, and T2). Delay sensitivity was expected to be improved in these programs; however, three actually displayed delay-induced errors in our phantom analysis. Delay sensitivity has been mainly focused on CBF and MTT (11,17); however, we have shown delay dependency for CBV, as well. Because CBV can be calculated by using the ratio of area under the curve between tissue and large vessel (16), delay-induced truncation of the tissue curve can affect the area under the curve and then CBV.

In addition to these delay sensitivities, other sources of errors, such as true CBV or true MTT dependencies, were evaluated by using our digital phantom. CBF and MTT calculations should ideally be independent of CBV values, and conversely, the CBV calculation should not be affected by MTT values. However, these dependencies were observed in all the algorithms used herein for at least one of the three parameters both for CT perfusion and perfusion-weighted imaging. To increase the accuracy of perfusion analysis, these dependencies should be improved. Further, the shape of $R(t)$ might have a significant effect on quantification of hemodynamic parameters, as $R(t)$ dependencies were found in almost all the algorithms.

In regard to absolute values, slope and intercept in linear regression analysis were found to be significantly variable, suggesting that absolute values cannot be directly compared among different software programs. In addition, we cannot reliably determine the penumbra through threshold levels, nor can we compare any quantitative information between studies conducted by using absolute values. Variations in the absolute values may be partly caused by the difference in basic algorithms, such as SVD and maximum slope, as well as other minor implementations in perfusion analysis. Further, a difference in the intercept causes errors not only in absolute but also in relative values, especially for perfusion-weighted imaging.

Standardization of perfusion analysis is important for multicenter trials. If different programs produce different perfusion results, perfusion images cannot be used for patient selection or estimation of penumbra and final infarction. On the other hand, if standardization is achieved, the results obtained in one study with the use of one software package can be applied to another study with the use of a different program. In the present study, T_{max} values of most programs were in good agreement with the values from the Echoplanar Imaging Thrombolysis Evaluation Trial (19), and this fact may facilitate the use of T_{max} in large clinical trials in the future, where a number of perfusion analysis programs could be used. T_{max} is not a physiologic parameter such as CBF, CBV, and MTT, but researchers

(21) have shown that it closely relates to delay, dispersion, and MTT. Validation of perfusion programs by using digital phantom data may also facilitate the development of a more accurate and robust analysis and finally make standardized quality control a reality.

The present study had a number of limitations. First, three types of $R(t)$ were used: exponential, linear, and box shaped. Another shape of $R(t)$ may be used for this simulation (22); however, the true shape of $R(t)$ in the human brain is unknown, and these three types of $R(t)$ are traditionally used in tracer kinetic simulation studies (16,17). Second, the curves were generated only for the first-pass bolus of contrast material, and the second pass was not introduced. In real perfusion data, small peaks are observed at second pass. If the second pass had been included in the simulation, the data may have been more realistic. However, interpretation of results may be more complicated, because the results would be affected by the difference in postprocessing during handling or compensating for second-pass peaks. Third, only the average values for each tile were evaluated. If the standard deviation within a tile had been measured, the difference in the denoising process of each program could have been evaluated. However, our aim was to evaluate not the denoising process but the accuracy and reliability of perfusion analysis. Further studies may be conducted to examine more factors of various parameters, such as noise level, temporal resolution, and $R(t)$, on the basis of our digital phantom scheme. Fourth, we defined accuracy of each algorithm by the correlation coefficient to true values. It has been reported that linearity was lost in the high-flow region in certain types of algorithms, especially for low signal-to-noise ratio (17). However, we used linear regression analysis, as it seemed to be suitable to measure the difference between true and measured values. Fifth, correction for multiple comparisons was not conducted in the analysis of error sources. If the correction is used, statistical comparison may become stricter. However, as the aim of this study was a comparison among different algorithms, a simple and uncorrected comparison was performed.

In conclusion, a digital phantom is a useful tool for evaluating the accuracy and characteristics of CT perfusion and perfusion-weighted imaging perfusion analysis software. Our phantom analysis revealed that all the commercial programs for CT perfusion and perfusion-weighted imaging studied here showed delay-induced errors and/or insufficient correlations with true values, while all the academic programs for perfusion-weighted imaging studied here showed better correlations with true values.

Acknowledgments: The authors thank K. Segawa and H. Kabasawa from GE Healthcare, O. Miyazaki and T. Takahashi from Hitachi Medical Corporation, S. Kuwahara and Y. Suzuki from Philips Electronics Japan, T. Itoh and M. Nielsen from Siemens Healthcare, and Y. Ikeda from Toshiba Medical Systems for their help in the perfusion data analysis.

Disclosures of Conflicts of Interest: K.K. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: was paid consultancy fee by Medical Image Lab and received payment for lectures including service on speakers bureaus by GE Healthcare and Hitachi Medical Corporation. Other relationships: none to disclose. **S.C.** No relevant conflicts of interest to disclose. **M.S.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: author was paid consultancy fee by Lundbeck; institution was paid consultancy fee by Hitachi Medical; was paid for lectures including service on speakers bureaus by GE Healthcare, Otsuka Pharma, Mitsubishi Pharma, Daiichi Pharma, J-Mac System, Hitachi Medical, Olea Medical, Bayer, Sanofi, and Esai. Other relationships: none to disclose. **L.Ø.** No relevant conflicts of interest to disclose. **H.S.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: paid for board membership by the Japanese Organization of Radiotherapy Quality Management and for consultancy by Medical Image Lab; institution received grants from Hitachi, Hitachi Medical, Varian Medical Systems, Shimadzu, Hokkaido Electric Power, and Elekta K.K.; institution received collaborative research grants from Accuthera, Tanaka Kikinzoku Kogyo K.K., Medikit, Mitsubishi Heavy Industries, Philips Electronics Japan, Fujifilm Medical, Sumitomo Heavy Industries, Shimadzu, Asahikawa Medical University, Nakashima Medical, Hokkaido Gastrointestinal Cancer Study group; institution received consigned research grants from Fujifilm Medical; received payment for lectures including service on speakers bureaus from Moroo, Hitachi, Chiyoda Technol, Japan Society of Medical Physics, Olympus, Varian Medical Systems, and ORIX Life Insurance. Other relationships: none to disclose. **K.O.** No relevant conflicts of interest to disclose. **M.W.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: institution received a grant from GE Healthcare. Other relationships: none to disclose. **S.W.** No relevant conflicts of interest to disclose.

References

1. Klotz E, König M. Perfusion measurements of the brain: using dynamic CT for the quantitative assessment of cerebral ischemia in acute stroke. Eur J Radiol 1999;30(3):170–184.
2. Mayer TE, Hamann GF, Baranczyk J, et al. Dynamic CT perfusion imaging of acute stroke. AJNR Am J Neuroradiol 2000;21(8):1441–1449.
3. Lev MH, Segal AZ, Farkas J, et al. Utility of perfusion-weighted CT imaging in acute middle cerebral artery stroke treated with intra-arterial thrombolysis: prediction of final infarct volume and clinical outcome. Stroke 2001;32(9):2021–2028.
4. Jain R, Hoeffner EG, Deveikis JP, Harrigan MR, Thompson BG, Mukherji SK. Carotid perfusion CT with balloon occlusion and acetazolamide challenge test: feasibility. Radiology 2004;231(3):906–913.
5. Laslo AM, Eastwood JD, Chen FX, Lee TY. Dynamic CT perfusion imaging in subarachnoid hemorrhage-related vasospasm. AJNR Am J Neuroradiol 2006;27(3):624–631.
6. Cenic A, Nabavi DG, Craen RA, Gelb AW, Lee TY. A CT method to measure hemodynamics in brain tumors: validation and application of cerebral blood flow maps. AJNR Am J Neuroradiol 2000;21(3):462–470.
7. Olivot JM, Mlynash M, Thijs VN, et al. Optimal Tmax threshold for predicting penumbral tissue in acute stroke. Stroke 2009;40(2):469–475.
8. Davis SM, Donnan GA. Advances in penumbra imaging with MR. Cerebrovasc Dis 2004;17(Suppl 3):23–27.
9. Rose SE, Janke AL, Griffin M, Finnigan S, Chalk JB. Improved prediction of final infarct volume using bolus delay-corrected perfusion-weighted MRI: implications for the ischemic penumbra. Stroke 2004;35(11):2466–2471.
10. Sasaki M, Kudo K, Honjo K, Hu JQ, Wang HB, Shintaku K. Prediction of infarct volume and neurologic outcome by using automated multiparametric perfusion-weighted magnetic resonance imaging in a primate model of permanent middle cerebral artery occlusion. J Cereb Blood Flow Metab 2011;31(2):448–456.
11. Kudo K, Sasaki M, Ogasawara K, Terae S, Ehara S, Shirato H. Difference in tracer

delay-induced effect among deconvolution algorithms in CT perfusion analysis: quantitative evaluation with digital phantoms. Radiology 2009;251(1):241–249.

12. Kudo K, Sasaki M, Yamada K, et al. Differences in CT perfusion maps generated by different commercial software: quantitative analysis by using identical source data of acute stroke patients. Radiology 2010; 254(1):200–209.
13. Christensen S, Mouridsen K, Wu O, et al. Comparison of 10 perfusion MRI parameters in 97 sub-6-hour stroke patients using voxel-based receiver operating characteristics analysis. Stroke 2009;40(6):2055–2061.
14. Dani KA, Thomas RG, Chappell FM, et al. Computed tomography and magnetic resonance perfusion imaging in ischemic stroke: definitions and thresholds. Ann Neurol 2011; 70(3):384–401.
15. Wintermark M, Albers GW, Alexandrov AV, et al. Acute stroke imaging research roadmap. Stroke 2008;39(5):1621–1628.
16. Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. I. Mathematical approach and statistical analysis. Magn Reson Med 1996;36(5):715–725.
17. Wu O, Østergaard L, Weisskoff RM, Benner T, Rosen BR, Sorensen AG. Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix. Magn Reson Med 2003;50(1):164–174.
18. Meier P, Zierler KL. On the theory of the indicator-dilution method for measurement of blood flow and volume. J Appl Physiol 1954; 6(12):731–744.
19. Davis SM, Donnan GA, Parsons MW, et al. Effects of alteplase beyond 3 h after stroke in the Echoplanar Imaging Thrombolytic Evaluation Trial (EPITHET): a placebo-controlled randomised trial. Lancet Neurol 2008; 7(4):299–309.
20. Straka M, Albers GW, Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke. J Magn Reson Imaging 2010; 32(5):1024–1037.
21. Calamante F, Christensen S, Desmond PM, Ostergaard L, Davis SM, Connelly A. The physiological significance of the time-to-maximum (Tmax) parameter in perfusion MRI. Stroke 2010;41(6):1169–1174.
22. Mouridsen K, Friston K, Hjort N, Gyldensted L, Østergaard L, Kiebel S. Bayesian estimation of cerebral perfusion using a physiological model of microvasculature. Neuroimage 2006;33(2):570–579.