Observer Variability in CT Perfusion Parameters in Primary and Metastatic Tumors in the Lung

Chaan S. Ng, MD, PhD[1], Wei Wei, MS[2], Payel Ghosh, PhD[1],
Ella Anderson, RT[1], Delise H. Herron, RT[1], and Adam G. Chandler, PhD[3]

Abstract

Purpose: Evaluate observer variability in computed tomography perfusion measurements in lung tumors and assess the relative contributions of individual factors to overall variability. **Materials and Methods:** Four observers independently delineated tumor and defined arterial input function region of interests (tumor region of interest and arterial input function region of interest) on each of 4 contiguous slice levels of computed tomography perfusion images (arterial input function level), in 12 computed tomography perfusion data sets containing lung tumors (>2.5 cm size), on 2 separate occasions. Computed tomography perfusion parameters (blood flow, blood volume, mean transit time, and permeability surface area product) for tumor volumes of interest were computed for all combinations of these factors, totaling up to 1024 combinations per patient. Overall, inter- and intraobserver variability were assessed by within-patient coefficient of variation, variance components analyses, and intraclass correlation. **Results:** Overall observer within-patient coefficient of variations for tumor blood flow, blood volume, mean transit time, and permeability surface area product were 20.3%, 11.9%, 6.3%, and 31.7%, and intraclass correlations were 0.94, 0.91, 0.82, and 0.72, respectively. Interobserver tumor volume of interest and arterial input function level were the highest contributors to overall variance for blood flow, blood volume, and mean transit time. Overall intraobserver wCVs for blood flow, blood volume, mean transit time, and permeability surface area product (4.3%, 2.4%, 0.9%, and 3.1%) were smaller than inter-observer within-patient coefficient of variations (9.5%, 5.6%, 1.6%, and 7.0%), respectively. **Conclusion:** The largest contributors to observer variability were interobserver tumor volume of interest and arterial input function level. Overall variability in computed tomography perfusion studies can potentially be minimized by using a single observer and a consistent level for arterial input function, which would be important considerations in longitudinal and multicenter studies. Methods to reliably define arterial input function and delineate tumor volumes would help to reduce variability in estimations of computed tomography perfusion parameter values.

Keywords

CT perfusion, lung tumors, observer variability

Abbreviations

AIF, arterial input function; AIF_{lev}, arterial input function level; AIF_{roi}, arterial input function region of interest; BF, blood flow; BV, blood volume; CT, computed tomography; CTp, computed tomography perfusion; ICC, intraclass correlation; IRB, institutional review board; MTT, mean transit time; NaN, not a number; PS, permeability surface area product; ROI, region of interest; TUM_{roi}, tumor region of interest; TUM_{voi}, tumor volume of interest; wCV, within-patient coefficient of variation.

[1] Department of Radiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA
[2] Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX, USA
[3] CT Research, GE Healthcare, Waukesha, WI, USA

Corresponding Author:
Chaan S. Ng, MD, PhD, Department of Radiology, Unit 1473, University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA.
Email: cng@mdanderson.org

Introduction

Computed tomography perfusion (CTp) is an evolving technique that is able to generate parameter values describing the perfusion characteristics of tissues, such as blood flow (BF), blood volume (BV), mean transit time (MTT), and permeability surface area product (PS).[1-3] An understanding of the variability, or reproducibility, associated with the measurements of CTp variables, and the factors affecting it, are important for its potential utility in clinical and quantitative assessments. Observer variability is an important potential contributor to overall variability of CTp measurements.

The computation of body CTp parameters are dependent essentially on 2 user-defined factors: delineation of the target lesion or tissue under consideration, with the drawing of a region of interest (ROI) around the lesion (tumor region of interest, [TUM$_{roi}$]) and definition of an arterial input function (AIF$_{roi}$). Both lead to respective time–intensity curves from which CTp parameters are derived with appropriate physiological modeling, in the work discussed herein, the distributed parameter model.[1] Computed tomography perfusion data sets typically consist of contiguous transaxial imaging slices, and these 2 fundamental determinants of CTp parameter values can in principle be defined on individual transaxial levels. In the case of tumor ROI delineation, this is typically undertaken on each slice location in which the tumor is visualized. In the case of AIF definition, this in principle could be obtained on any slice level (AIF$_{lev}$); little is known about the practical impact on resultant CTp parameter values of utilizing different slice locations for the definition of this input parameter.

Previous evaluations of observer variability in CTp have been undertaken in the brain[4-6] and for tumors in a variety of body locations.[7-11] The lung is a major site for primary and metastatic tumors.[12] Unlike evaluations in the brain, lesions in body locations are susceptible to motion (breathing, cardiac movement, etc), which imposes additional challenges in delineating ROIs. There has been little work on assessing the interaction of tumor ROI delineation and AIF definition, specifically in body tumors.[13] In addition, to the best of our knowledge, there have been no specific investigations evaluating the relative contributions of each of these fundamental factors to overall observer variability. The latter would provide insights as to the factors which impact most on variability.

Our primary objectives were to assess the observer variability of CT perfusion measurements derived from tumors in the lung and to assess the relative independent contributions to overall variability of observer-defined factors, including the relative impact of tumor ROI delineation, AIF definition, and between- and within-observer variability.

Methods and Materials

Patients

This retrospective study was approved by The University of Texas MD Anderson Cancer Center institutional review board (IRB), with waiver of informed consent. Data in this study were drawn from a previous prospective IRB-approved study which has been previously described.[14] In brief, the study consisted of adult patients who had lung tumors larger than 2.5 cm in longest axial diameter and had undergone 2 CTp scans, 2 to 7 days apart.

Computed Tomography Perfusion Scanning Technique

The CT perfusion scanning technique has been presented previously.[14] In brief, CTp scans were obtained in 2 phases: phase 1, cine acquisition during a 30 second breath-hold, followed 20 seconds later by phase 2, which consisted of 6 short breath-hold helical scans acquired at 15-second intervals.

Phase 1 (cine) scans were performed using a single level of 20-mm thickness (4i mode, 5-mm contiguous slice thickness, for 4 slices) at the midpoint of the target lesion, which had been selected on review of previous imaging studies by a radiologist (C.S.N., more than 15 years of experience in interpreting CT studies). Computed tomography data were collected at that single location using the cine mode for a 30-second breath-hold duration, with the following parameters: tube voltage, 120 kVp; tube current, 90 mA; rotation speed, 1 second; field of view, 34 to 42 cm; matrix, 512 × 512 mm (16-row multidetector CT scanner [LightSpeed, GE Healthcare, Waukesha, Wisconsin]). Data acquisition started 5 seconds after intravenous injection of 70 mL of a nonionic contrast agent (ioversol [Optiray], 320 mg of iodine/100 mL, Mallinckrodt Inc., St Louis, Missouri) using an automatic injector (MCT/MCT Plus; Medrad, Pittsburgh, Pennsylvania) and an injection rate of 7 mL/s. The phase 1 cine images were reconstructed to a temporal sampling interval of 0.5 second.

Phase 2 (delayed) scans consisted of 6 short intermittent helical scans, each obtained during a breath-hold of approximately 7 seconds, with 10-cm craniocaudal coverage centered on the target lesion, using the following parameters: tube voltage, 120 kVp; tube current, 90 mA; slice thickness, 5 mm; interval, 5 mm; pitch factor, 1.35; rotation speed, 0.8 s; field of view, 34 to 42 cm; and matrix, 512 × 512 mm. The final helical acquisition commencing 125 seconds after the start of the phase 1 acquisition. The helical volumes were reconstructed at 5-mm intervals.

Computed Tomography Perfusion Analysis and Observers

One CTp data set from each pair of CTs was acquired in each patient in the above prospective study, the one which was least affected by motion was used in the current analysis. Each data set was evaluated by 4 observers independently (C.S.N, E.F.A, D.H.H, A.G.C, each with more than 5 years of experience with CT perfusion analysis), on 2 occasions (rounds I and II) separated by more than 4 weeks, as described below. The acquired perfusion images were analyzed using commercially available CT perfusion software on a workstation (CT Perfusion version 4, Advantage Workstation 4.5; GE Healthcare).
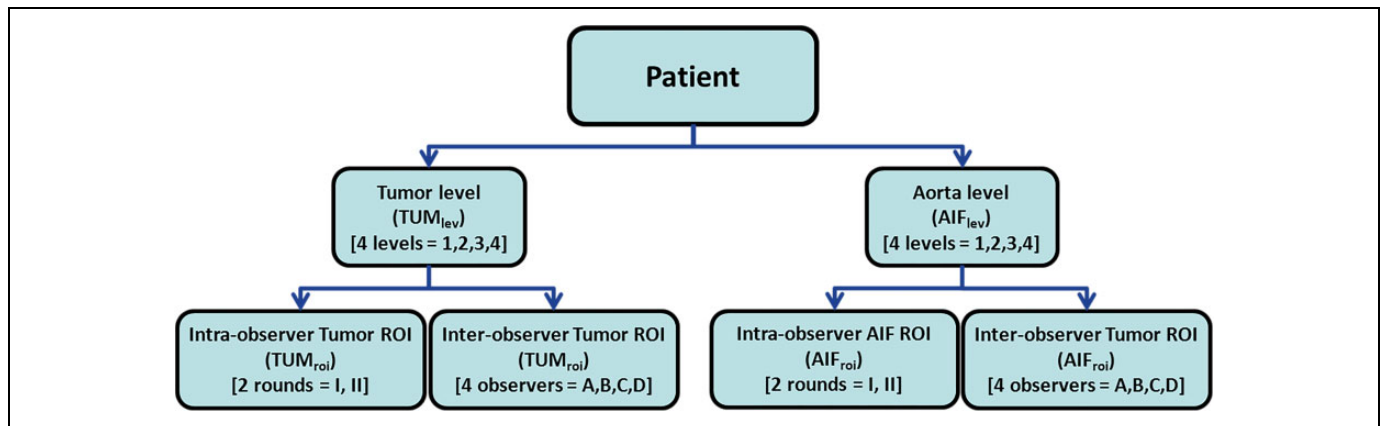
Figure 1. Schema of combination of analyses undertaken. Computed tomography image data consisted of 4 transaxial slices ("levels" 1, 2, 3, 4). There were 4 observers, each of whom undertook CT perfusion analyses (delineating tumor ROIs on up to 4 levels, and defining AIFs on each of 4 levels), on 2 separate occasions (rounds). As noted in the Methods, when considering all combinations of TUM_{roi} (4 AIF levels $\times$ 4 observers $\times$ 2 rounds) and AIF_{roi} (4 AIF levels $\times$ 4 observers $\times$ 2 rounds), this is total of 1024 combinations for each patient. AIF, arterial input function; CT indicates computed tomography; ROI, region of interest; TUM_{roi}, tumor region of interest.

Before the perfusion analyses were undertaken, motion correction was applied to the source CT images to limit the effects of breathing related misregistrations. Details of the registration technique have been described previously.[15] In brief, this postprocessing consisted of a semiautomated rigid registration algorithm applied to phase 1 and phase 2 images. The resultant image data set consisted of registered phase 1 images and 6 registered phase 2 images; this combined registered data set was used in subsequent analyses by all observers.

In our study, we used a commercially available CT perfusion software package to derive our tumor perfusion parameters (CT Perfusion version 4, Advantage Workstation 4.5; GE Healthcare). This package utilizes the distributed parameter physiological model of tissue perfusion, in which vascular/arterial input and tissue/tumor uptake functions are deconvolved to yield CTp parameters.[16]

Four CTp parametric maps (BF, BV, MTT, and PS) were generated as described above for the target lesion (tumor) at each of the 4 slices of the registered data set. For computation of the tumor CTp values, the 4 observers were required to: (a) delineate the region of interest of the target lesion (TUM_{roi}) and (b) define the arterial input function (AIF_{roi}). This is illustrated schematically in the first 2 rows of Figure 1.

For TUM_{roi} delineation, observers were presented with a paper printout of the target lesion of interest as identified at initial patient enrollment (this was simply to ensure that all observers used the same lesion for their CTp evaluations). The registered data sets were loaded into the Body Perfusion protocol of the above CT perfusion software. Window widths and levels were fixed and identical for all observers (W = 350, L = 40). Observers had freedom to scroll/cine through the data sets and then to delineate freehand target lesion ROIs using an electronic cursor and mouse on each of the CT levels in which tumor was visualized (Figure 2A).

For AIF definition, observers again loaded the data sets into the above CT perfusion software. Arterial input function definition is determined by both the specific AIF time–intensity curve and identification of the associated pre- and postenhancement "set points." The pre-enhancement set point defines the image/time at which arterial enhancement is considered to commence. The pre-enhancement set point and the profile of the AIF time–intensity curve are potentially affected by delineation of the ROI within the artery (AIF_{roi}) and hence the potential for observer variability. The postenhancement set point was the same in all cases (ie, the last acquired phase 2 time point). In this study, observers had freedom to place a round or oval ROI within the aorta and to adjust the location, size, and shape of the ROI to optimize the resultant AIF time–intensity curve (Figure 2B). General guidelines for aortic ROI delineation were that the resultant AIF time–intensity curve should have the highest peak Hounsfield attenuation possible and also minimize artifacts and noise. Observers recorded their associated "pre-enhancement" set points using the resultant AIF that their arterial ROIs generated (Figure 2B). Definition of the AIF was undertaken on each of the 4 levels of source data (AIF_{lev}).

All TUM_{roi}'s and AIF_{roi}'s generated by all observers, for all slice levels and observer rounds were saved for subsequent analyses. Tumor CTp parameters (BF, BV, MTT, and PS) were calculated using these saved TUM_{roi} and AIF_{roi}, for each imaging slice level of tumor using the AIFs defined on each level (AIF_{lev}) by the 4 different observers on their 2 separate rounds of observations (Figure 3). Computed tomography perfusion analyses were undertaken for all combinations of TUM_{roi} (4 AIF levels $\times$ 4 observers $\times$ 2 rounds) and AIF_{roi} (4 AIF levels $\times$ 4 observers $\times$ 2 rounds), namely, a possible total of 1024 combinations for each patient (Figure 1).

Pixel-level data for each TUM_{roi} from each level on which a tumor ROI was delineated was extracted and combined to form a single tumor volume of interest (TUM_{voi}), which formed the basis for all subsequent analyses. It was noted that the software labeled some pixels as "not a number" (NaN), in which case these pixels were excluded from all other CTp parameter maps
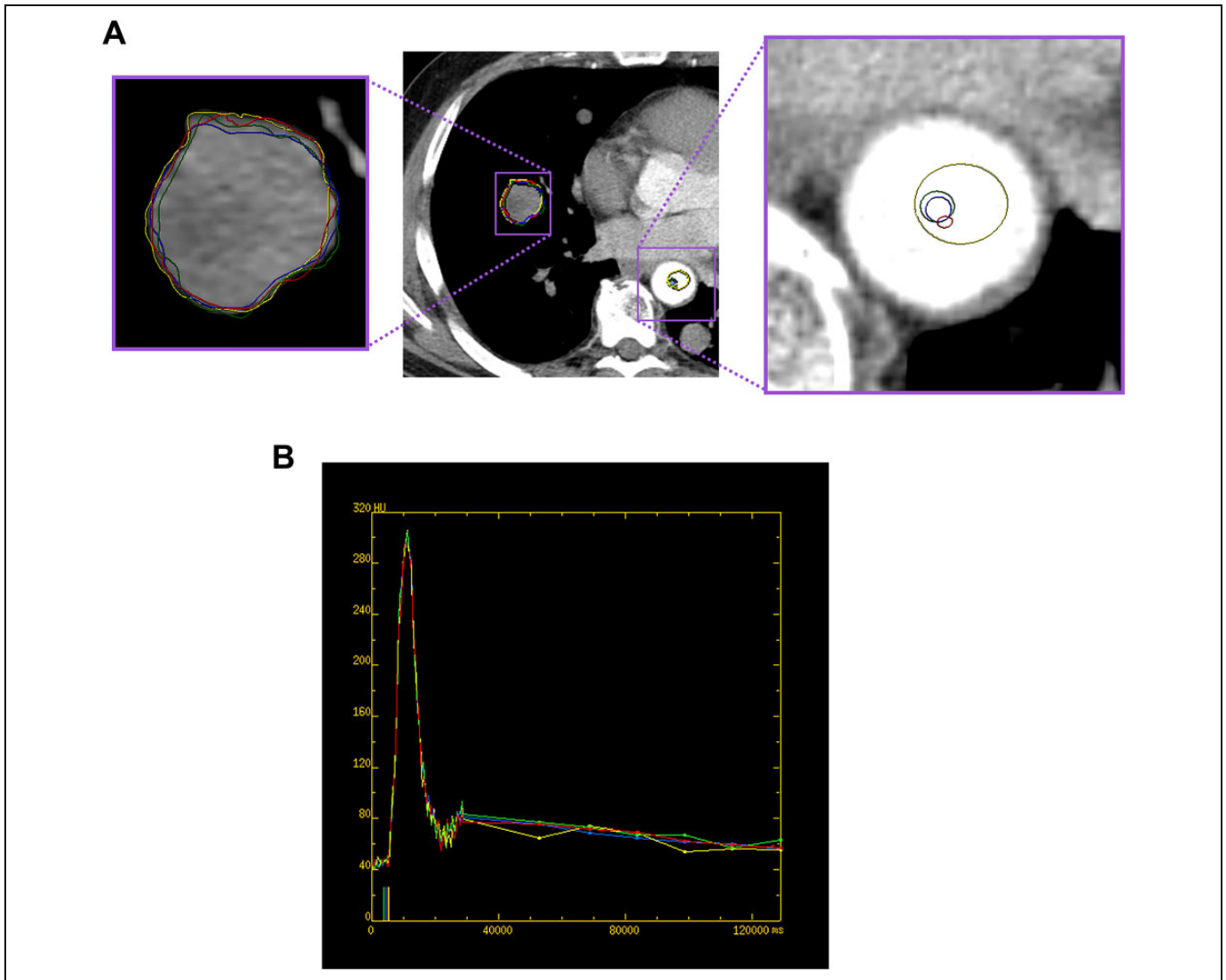
Figure 2. A 51-year-old male with metastatic right lung nodule from a spindle cell sarcoma. (A) Representative example of the TUM_{roi} and AIF_{roi} defined by the 4 observers (yellow, blue, green, and red ROIs). (B) Arterial input and tumor time–density curves for the AIF_{roi} and TUM_{roi} delineated by the 4 observers' ROI in (A; yellow, blue, green, and red lines), and the associated pre-enhancement set points chosen by the 4 observers (corresponding colored marks on x axis). y axis, Hounsfield units; x axis, time in milliseconds. AIF_{roi} indicates arterial input function region of interest; ROI, region of interest; TUM_{roi}, tumor region of interest.

for the same lesion. Similarly, we excluded any pixels in which PS > BF since such pixels would not be physiological.

Statistics

Distribution of raw measurements of the CTp parameters (BF, BV, MTT, and PS) and ROI sizes were examined by mean, standard deviations, median, and range. All CTp parameters were transformed to the cubic root scale prior to statistical analyses due to skewness in pixel distributions.

Computed tomography perfusion parameter values are computed fundamentally from the time–intensity curves associated with the ROIs drawn for tumor and arterial input, that is, TUM_{roi} and AIF_{roi}. Since observers inevitably delineate their ROIs differently on separate occasions and between

themselves, each of these has potential for intraobserver and interobserver components. Computed tomography perfusion values were obtained after combining pixel data from the delineated TUM_{roi}'s, that is, for the TUM_{voi}. There were 3 principle factors of interest: TUM_{voi}, AIF_{roi}, and AIF_{lev}, the first 2 of which have inter- and intraobserver components.

A variance component analysis was performed to estimate the individual sources of variations including between–aorta level variation (AIF_{lev}), inter-/intraobserver TUM_{voi}, and inter-/intraobserver AIF_{roi} variation for each CTp parameter, as well as overall between- and within-patient variation.[17] After each variance component analysis, the within-patient coefficient of variation (wCV) was calculated for each source of variation using the Bland-Altman method.[18]
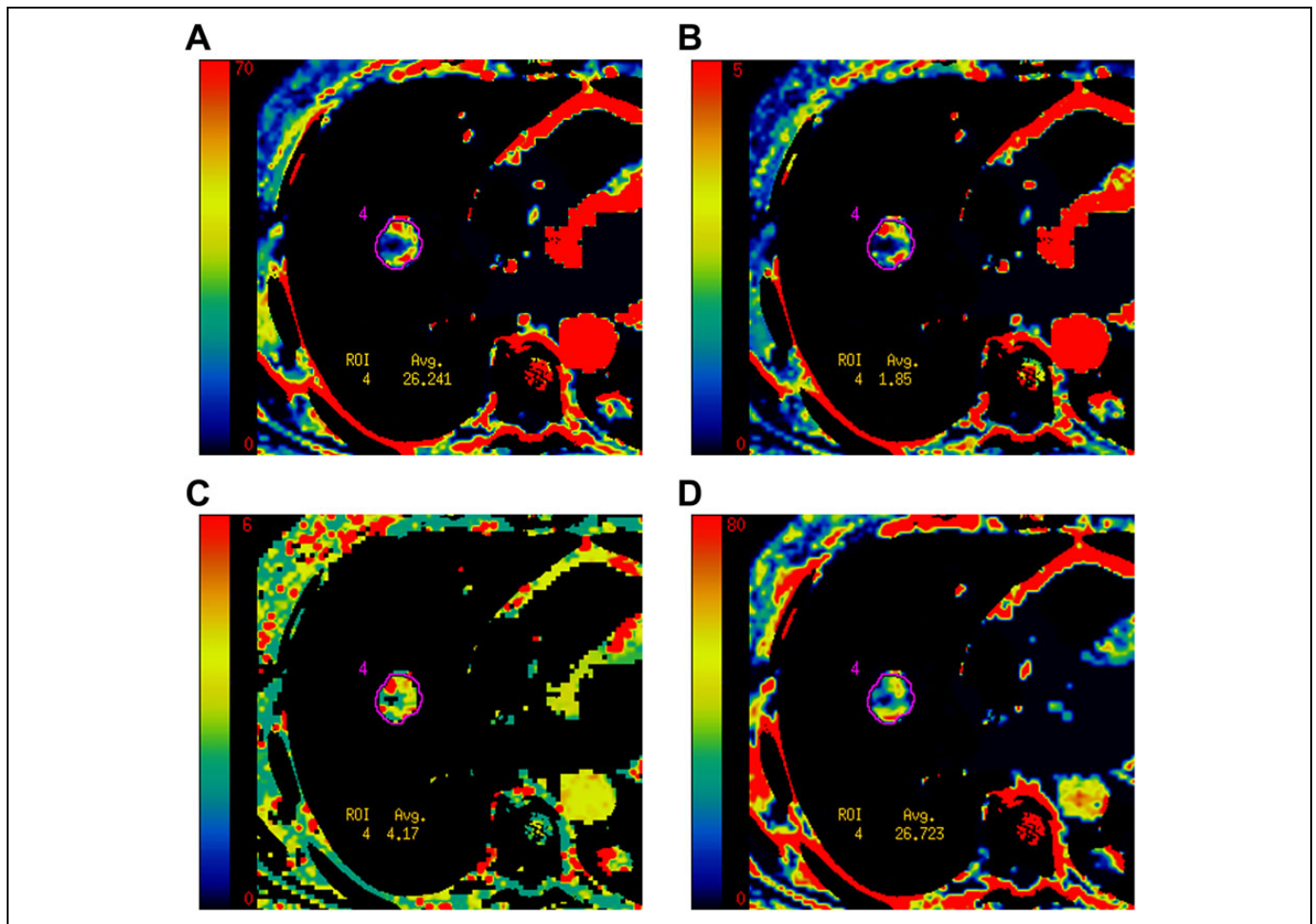
Figure 3. Representative CTp parametric maps for one of the observer combinations of TUM$_{roi}$ and AIF$_{roi}$ at one of the 4 CT slice levels. (A) BF, (B) BV, (C) MTT, and (D) PS. Same patient as in Figure 2. AIF$_{roi}$ indicates arterial input function region of interest; BE, blood flow; BF, blood flow; BV, blood volume; CTp, computed tomography perfusion; MTT, mean transit time; PS, permeability surface area product; TUM$_{roi}$, tumor region of interest.

Intraclass correlation (ICC) coefficients and their 95% confidence intervals for overall reproducibility for each CTp parameter were calculated using the 1-way random effect model (ICC[1]) described by McGraw and Wong.[19] Statistical analysis was carried out using SAS version 9.4 (SAS Institute, Cary, North Carolina).

Within-patient coefficient of variation and ICC provide complementary information. Within-patient coefficient of variation provides an estimate of individual source of variation, relative to the mean of the overall patient data. Intraclass correlation provides an assessment of overall within-patient similarity; in this all sources of variation are combined into 2 groups, within-patient and between-patient, and compared. A highly reproducible test would have wCV close to 0 and an ICC close to unity.

Results

Twelve patients contributed to the analysis (mean age 55.1 [range, 21.6-74.8 years] years; 8 male, 4 female). There were 5 patients with primary lung tumors and 7 with metastatic tumors to the lungs. The primary sites of malignancy of the 7 patients with secondary tumors were melanoma (n = 3), sarcoma (n = 2), renal (n = 1), and rectal (n = 1; Table 1).

The median size of the lung tumors was 4.5 cm (range, 2.5-9.7 cm). A total of 368 TUM$_{roi}$ were drawn by all 4 observers on 2 occasions, on up to 4 CT slice levels in which tumor was identified. The median tumor volumes of the 12 lung tumors was 31.3 cm^3 (range, 9.8-163.1 cm^3). Overall, 27% of voxels were not evaluable because of vendor software NaNs, and a further 46% because of the criterion BF < PS.

Arterial input functions were defined on all 4 slice levels, except in 1 patient in which it could only be defined on 1 aortic level (because of beam hardening artifact on other levels). A total of 360 AIF$_{roi}$'s were drawn, with median cross-sectional area of 46.4 mm^2 (interquartile range, 18.3-91.6 mm^2).

Median BF, BV, MTT, and PS values for the TUM$_{voi}$ were 58.0 mL/min/100 g (range, 15.8-236.1 mL/min/100 g), 4.7 mL/100 g (range, 0.6-21.0 mL/100 g), 5.1 seconds (range, 2.3-8.2 seconds), and 24.9 mL/min/100 g (range,

Table 1. Summary of Patients and Lung Lesions.[a]

Age (years)	Gender	Primary Tumor	Lesion Size (cm)
44.2	M	Lung	3.2
49.6	M	Lung	4.7
60.2	F	Lung	7.7
65.9	M	Lung	4.3
74.8	F	Lung	3.5
50.7	M	Melanoma	5.3
56.1	M	Melanoma	2.5
59.3	F	Melanoma	9.7
21.6	M	Ewing's sarcoma	5.2
64.8	F	Leiomyosarcoma	3.6
51.7	M	Renal	5.3
60.4	M	Rectal	3.6

Abbreviations: F, female; M, male.
[a]Reprinted in part with permission from *American Journal of Roentgenology*.

0.3-94.0 mL/min/100 g), respectively. The overall within-patient wCVs for all observers and all observer combinations of the above tumor and aortic ROIs and levels for BF, BV, MTT, and PS were 20.3%, 11.9%, 6.3%, and 31.7%, respectively (Table 2).

Contributors to Overall Variability

Our variance components analysis indicates that the highest contributors to overall observer variance differ according to the CTp parameter (Table 3). For BF and BV, the highest contributor was interobserver TUM_{voi}, that is, delineation of tumor ROIs between observers, with wCVs of 7.8% and 4.7%, respectively. The next highest contributor was AIF_{lev}, with wCVs of 6.2% and 3.6%, respectively. Interobserver AIF_{roi}, intraobserver TUM_{voi}, and intraobserver AIF_{roi} made successively smaller contributions to the overall observer variance, respectively.

For MTT, the 2 main contributors to overall observer variance were the same as for BF and BV, except in the reverse order, namely, AIF_{lev} (wCV, 1.8%) and TUM_{roi} (wCV, 1.5%), respectively. For PS, the main contributors to overall observer variance were related to the AIF, namely AIF_{lev}, interobserver AIF_{roi}, and intraobserver AIF_{roi}, with wCVs of 8.7%, 7.0%, and 3.1%, respectively.

Interobserver variability (wCV, 1.6%-9.5%) was consistently higher than corresponding intraobserver variability (wCV, 0.9%-4.3%) across all 4 CTp parameters (Table 2). No systematic trends in perfusion values with AIF_{lev} were evident.

Intraclass Correlation

The overall ICCs for all observers and all observer combinations of tumor and aortic ROIs and levels for BF, BV, MTT, and PS were 0.94, 0.91, 0.81, and 0.72, respectively (Table 4).

Observer Variability Using One Aortic Level

Given our findings above that AIF_{lev} is one of the main contributors to observer variability, we undertook a supplementary analysis to examine the impact on overall observer variability in CTp values of constraining AIF definition to one consistent, fixed AIF_{lev}. The within-patient wCVs for all observers and all observer combinations of tumor and aortic ROIs, with a consistent fixed aortic level, for BF, BV, MTT, and PS were 16.2%, 9.4%, 5.4%, and 24.1%, respectively (Table 2). These wCVs are lower than for the above wCVs obtained for all observers and all aortic levels.

Discussion

An assessment of observer variability is an important component in evaluating the clinical utility of any test, particularly a quantitative one. In our study, we have attempted to provide an overview of the *overall* observer variability for CTp, that is, the variability that might be seen if data were processed by different observers. Our results yield overall observer wCVs for BF, BV, MTT, and PS of 20.3%, 11.9%, 6.3%, and 31.7%, respectively.

This overarching information is obscured in previous studies, which instead have typically compartmentalized their results into inter- and intraobserver variability. Our study examines these specific subcomponents of overall variability, but in addition attempts to dissect out and quantify the relative independent contributions to the overall observer variability of the individual user-defined factors, specifically, delineation of the TUM_{voi} and definition of the AIF_{roi}. In the process, we have also examined the contribution to overall variability of the slice level/location of the AIF_{lev}. Identification of the major source(s) of variability provides the opportunity to develop appropriate strategies to mitigate their impact.

Our results suggest that the main contributors to overall observer variability vary according to the CTp parameters. For BF and BV, the main source of variability was TUM_{voi}, with wCVs 7.8% and 4.7%, respectively, followed by AIF_{lev}, with wCVs of 6.2% and 3.4%, respectively. These 2 variables were also the main contributors to variability in MTT. Of note, AIF_{lev} contributed higher variance than interobserver and intraobserver AIF_{roi}, suggesting that the slice level on which AIFs are placed might contribute more to variability than other factors related to the definition of the AIF, such as size of ROI and definition of the pre-enhancement set points. A practical implication is that CTp variability would likely be reduced if analyses were performed with AIFs defined on 1 fixed and consistent level (such as the most superior CT slice) within longitudinal and/or multicenter studies.

Most previous studies have assessed observer variability in terms of the interobserver and intraobserver components of variability. In our study, the overall intraobserver variability was lower compared to interobserver variability, as has been noted in most other studies. Our intra- and interobserver wCVs for BF, BV, and MTT were in the range 0.9% to 4.3% and 1.6%

Table 2. Summary of CT Perfusion Parameter Values for All Patients and All Observations, and (a) Overall Within-Patient, Interobserver, and Intraobserver Variations With All AIF$_{lev}$, and (b) Within-Patient Variation With 1 Fixed AIF$_{lev}$.[a]

	Mean[c]	SD[c]	wCV (%) All AIF$_{lev}$			wCV (%) Single AIF$_{lev}$[b]
			Overall Within-Patient	Overall Interobserver	Overall Intraobserver	Overall Within-Patient
BF	4.1	0.7	20.3	9.5	4.3	16.2
BV	1.7	0.3	11.9	5.6	2.4	9.4
MTT	1.7	0.1	6.3	1.6	0.9	5.4
PS	2.9	0.5	31.7	7.0	3.1	24.1

Abbreviations: AIF$_{lev}$, arterial input function level; AIF$_{roi}$, arterial input function; BF, blood flow; BV, blood volume; CT, Computed tomography; MTT, mean transit time; PS, permeability surface area product; SD, standard deviation; TUM$_{voi}$, tumor volume of interest; wCV, within-patient coefficient of variation, percent.
[a]Interobserver wCV consists of evaluation of interobserver AIF$_{roi}$ and interobserver TUM$_{voi}$; intraobserver wCV consists of evaluation of intraobserver AIF$_{roi}$ and intraobserver TUM$_{voi}$.
[b]Most superior AIF$_{lev}$.
[c]Mean and SD on cubic root scale of CT perfusion parameter values for TUM$_{voi}$.

Table 3. Computed Tomography Perfusion Parameter Variances and wCVs, by Source of Variation (on Cubic Root Scale).

Parameter	Source of Variation	Estimated Variance	wCV (%)
BF	Between-patient	0.5067	
	AIF$_{lev}$ between–aorta level	0.003648	6.23
	AIF$_{roi}$ interobserver	0.002584	5.21
	AIF$_{roi}$ intraobserver	0.000045	0.68
	TUM$_{voi}$ interobserver	0.005657	7.81
	TUM$_{voi}$ intraobserver	0.001741	4.26
BV	Between-patient	0.1114	
	AIF$_{lev}$ between–aorta level	0.001235	3.58
	AIF$_{roi}$ interobserver	0.000784	2.84
	AIF$_{roi}$ intraobserver	0.000025	0.5
	TUM$_{voi}$ interobserver	0.002134	4.73
	TUM$_{voi}$ intraobserver	0.000556	2.39
MTT	Between-patient	0.01555	
	AIF$_{lev}$ between–aorta level	0.000316	1.79
	AIF$_{roi}$ interobserver	0.000012	0.34
	AIF$_{roi}$ intraobserver	0.00	0
	TUM$_{voi}$ interobserver	0.000233	1.54
	TUM$_{voi}$ intraobserver	0.000076	0.88
PS	Between-patient	0.1834	
	AIF$_{lev}$ between–aorta level	0.006917	8.67
	AIF$_{roi}$ interobserver	0.004525	6.96
	AIF$_{roi}$ intraobserver	0.000940	3.11
	TUM$_{voi}$ interobserver	0.000082	0.91
	TUM$_{voi}$ intraobserver	0.00	0

Abbreviations: AIF$_{lev}$, arterial input function level; AIF$_{roi}$, arterial input function region of interest; BF, blood flow; BV, blood volume; CT, computed tomography; MTT, mean transit time; PS, permeability surface area product; TUM$_{voi}$, tumor volume of interest; wCV, within-patient coefficient of variation, percent.

Table 4. Intraclass Correlation (ICC) Coefficients for All Observers and All Combinations of Observations.

Parameter	ICC	ICC 95%CI
BF	0.94	0.89-0.97
BV	0.91	0.83-0.95
MTT	0.82	0.69-0.90
PS	0.72	0.56-0.84

Abbreviations: BF, blood flow; BV, blood volume; CI, confidence interval; MTT, mean transit time; PS, permeability surface area product.

noted that wCVs are not simply additive. Nevertheless, a similar study of 12 patients with primary lung tumors and 2 observers reported intraobserver wCVs for BF, BV, and MTT of 1.36% to 3.03%, 1.21% to 2.86%, and 1.05% to 2.28%, respectively, and interobserver wCVs of 1.81% to 4.69%, 1.14% to 4.63%, and 1.21% to 1.34%, respectively.[10] A study of 20-lung CT perfusion data sets evaluated by 2 observers, using Patlak physiological modeling to compute tumor BV and permeability, reported intraobserver and interobserver wCVs of 3.30% to 6.34%.[9] The general finding that intraobserver variation is smaller compared to interobserver variability in our study and prior studies[4,5,7,9] suggests that the overall variability would probably be reduced if all measurements within a study were undertaken by a single observer. We recognize that this might not always be practical, but it is what our, and indeed other, data suggest.

Assessment by ICC indicated relative high correlations for the 4 CT perfusion parameters in our study (0.72-0.94), albeit lower than in the studies above, which reported ICCs of 0.97 to 0.99.[9,10] Part of this may be because we assessed ICC based on all observations and all observer combinations, while the other studies divided their ICCs by inter- and/or intraobserver components. In addition, our assessment of ICC incorporates the variations of 4 observers, in contrast to just 2 observers as in the other studies. It is probably of greater interest to have an overview of the ICC of all observers in a study, than of just the

to 9.5%, respectively. Comparisons with other studies should be undertaken with caution since there are differences in the physiological models utilized, CT acquisition techniques that have been studied, tissues evaluated, specifics of study design, and statistical methods employed. Furthermore, it should be

individual (within-observer) variabilities of each observer. It is probably of even more interest to have a summation of the overall variability across all observers and all factors that contribute to observer variability, as assessed in our study. As for our wCV analyses, our overall ICCs have pooled all our observer data, incorporating both inter- and intraobserver variations.

Our results suggest that there might be some differences in the observer variability across the 4 CTp parameters; specifically, it appears that higher observer variability might be associated with PS values than with the other 3 CTp parameters. These observations may be a reflection of higher intrinsic noise/error associated with the CTp estimations of PS. Our variance components analysis also suggests that for PS, AIF-related factors make a relatively larger contribution to variance than TUM_{voi} delineations, which is unlike BF, BV, and MTT. This may be because estimations of PS, in contrast to the other 3 parameters, is determined more by the later (rather than the earlier) portions of the time–intensity curves. It will be noted in Figure 2B that the later time points of the aortic input curve show somewhat more variability among the 4 observers than in the initial 30 seconds, which likely contributes to variability in curve fitting and resultant estimations of PS.

Strategies to improve observer variation, apart from limiting AIF_{lev} to a single level as discussed above, might include the utilization of automated or semiautomated methods to delineate/segment tumor ROIs and to define AIFs. The former might involve segmentation algorithms, and the latter, identification of optimized or "idealized" arterial inputs from vessels.[20] The development of accurate and robust segmentation algorithms is currently an intense area of research and is beyond the scope of this work. Limited reports in brain CTp suggest that (semi) automation of the postprocessing involved in CTp analyses may help to reduce observer variability.[6,21] Unlike CTp in the brain, motion in body/thoracic CTp introduces additional challenges, for example, delineation of tumor boundaries, for which high-quality motion correction will be important.

Unlike previous studies, we interrogated the pixel-level data of our TUM_{voi}. In so doing, we identified the presence of software-related NaNs, which is not in itself surprising. We also identified pixels in which PS > BF, which were excluded because of their nonphysiologic nature.

We acknowledge and recognize several limitations of our work. The number of patients in our study was relatively small; however, unlike other studies, it incorporated multiple observers (not just 2, as in most previous studies); it has also attempted to dissect out the major sources of observer variation. Our source data was limited to a 20-mm *z* axis acquisition; newer scanners have the potential for more extensive *z* axis coverage. Although this might have some impact on our results, it is unlikely to have substantial effects since all observers in our observer study evaluated the same data sets. It is quite possible that other tissue/tumor types, physiological modeling, and acquisition parameters may yield different results; such investigation was beyond the scope of the current work and would require separate evaluation.

In summary, both tumor ROI delineation and AIF definition contribute to observer variability in the estimation of CTp parameter values. Of the potential sources of observer variation, TUM_{voi} and AIF_{lev} contributed the most to overall observer variability for the majority of CT perfusion parameters. Intraobserver variability was smaller than for interobserver variability. Overall observer variability in CT perfusion measurements would be minimized by utilizing a fixed and consistent AIF_{lev} and a single observer, which would be important considerations particularly in longitudinal and multicenter studies. Methods to reliably define AIF and delineate tumor volumes would likely help to reduce variability in estimations of CT perfusion parameter values.

Declaration of Conflicting Interests

Funding

References

1. Lee TY, Purdie TG, Stewart E. CT imaging of angiogenesis. *Q J Nucl Med*. 2003;47(3):171-187.

2. Miles KA, Griffiths MR. Perfusion CT: a worthwhile enhancement? *Brit J Radiol*. 2003;76(904):220-231.

3. Kambadakone AR, Sahani DV. Body perfusion CT: technique, clinical applications, and advances. *Radiol Clin North Am*. 2009; 47(1):161-178.

4. Fiorella D, Heiserman J, Prenger E, Partovi S. Assessment of the reproducibility of postprocessing dynamic CT perfusion data. *AJNR*. 2004;25(1):97-107.

5. Waaijer A, van der Schaaf IC, Velthuis BK, et al. Reproducibility of quantitative CT brain perfusion measurements in patients with symptomatic unilateral carotid artery stenosis. *AJNR*. 2007;28(5): 927-932.

6. Soares BP, Dankbaar JW, Bredno J, et al. Automated versus manual post-processing of perfusion-CT data in patients with acute cerebral ischemia: influence on interobserver variability. *Neuroradiology*. 2009;51(7):445-451.

7. Goh V, Halligan S, Hugill JA, Bassett P, Bartram CI. Quantitative assessment of colorectal cancer perfusion using MDCT: inter- and intraobserver agreement. *AJR Am J Roentgenol*. 2005;185(1): 225-231.

8. Goh V, Halligan S, Hugill JA, Bartram CI. Quantitative assessment of tissue perfusion using MDCT: comparison of colorectal cancer and skeletal muscle measurement reproducibility. *AJR Am J Roentgenol*. 2006;187(1):164-169.

9. Ng QS, Goh V, Fichte H, et al. Lung cancer perfusion at multi-detector row CT: reproducibility of whole tumor quantitative measurements. *Radiology*. 2006;239(2):547-553.

10. Larici AR, Calandriello L, Amato M, et al. First-pass perfusion of non-small-cell lung cancer (NSCLC) with 64-detector-row CT: a study of technique repeatability and intra- and interobserver variability. *Radiol Med*. 2014;119(1):4-12.

11. Ippolito D, Casiraghi AS, Talei Franzesi C, Bonaffini PA, Fior D, Sironi S. Intraobserver and interobserver agreement in the evaluation of tumor vascularization with computed tomography perfusion in cirrhotic patients with hepatocellular carcinoma. *J Comput Assist Tomogr*. 2016;40(1):152-159.

12. Husband JE, Reznek RH. *Husband and Reznek's Imaging in Oncology*. 3rd ed. London: Informa Healthcare; 2010.

13. Sanelli PC, Lev MH, Eastwood JD, Gonzalez RG, Lee TY. The effect of varying user-selected input parameters on quantitative values in CT perfusion maps. *Acad Radiol*. 2004;11(10):1085-1092.

14. Ng CS, Chandler AG, Wei W, et al. Reproducibility of perfusion parameters obtained from perfusion CT in lung tumors. *AJR Am J Roentgenol*. 2011;197(1):113-121.

15. Chandler A, Wei W, Herron DH, Anderson EF, Johnson VE, Ng CS. Semiautomated motion correction of tumors in lung CT-perfusion studies. *Acad Radiol*. 2011;18(3): 286-293.

16. Lee TY. Functional CT: physiological models. *Trend Biotechnol*. 2002;20(8):S3-S10.

17. Rawlings JO, Pantula SG, Dickey DA. *Applied Regression Analysis: A Research Tool*. 2nd ed. New York: Springer Science & Business Media; 1998.

18. Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ*. 1996;313(7049):106.

19. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Meth*. 1996;1(1):30-46.

20. Ashton E, Raunig D, Ng C, Kelcz F, McShane T, Evelhoch J. Scan-rescan variability in perfusion assessment of tumors in MRI using both model and data-derived arterial input functions. *J Magn Reson Imaging*. 2008;28(3):791-796.

21. Sanelli PC, Nicola G, Tsiouris AJ, et al. Reproducibility of postprocessing of quantitative CT perfusion maps. *AJR Am J Roentgenol*. 2007;188(1):213-218.