Summation or Axial Slab Average Intensity Projection of Abdominal Thin-section CT Datasets: Can They Substitute for the Primary Reconstruction from Raw Projection Data?

Kyoung Ho Lee,^{1,2} Helen Hong,³ Seokyoung Hahn,^{4,5} Bohyoung Kim,^{1,2} Kil Joong Kim,^{1,2} and Young Hoon Kim^{1,2}

We hypothesized that the summation or axial slab average intensity projection (AIP) techniques can substitute for the primary reconstruction (PR) from a raw projection data for abdominal applications. To compare with PR datasets (5-mm thick, 20% overlap) in 150 abdominal studies, corresponding summation and AIP datasets were calculated from 2-mm thick images (50% overlap). The root-mean-square error between PR and summation images was significantly greater than that between PR and AIP images (9.55 [median] vs. 7.12, $p < 0.0001$, Wilcoxon signed-ranks test). Four radiologists independently compared 2,000 test images (PR [as control], summation, or AIP) and their corresponding PR images to prove that the identicalness of summation or AIP images to PR images was not 1% less than the assessed identicalness of PR images to themselves (Wald-type test for clustered matched-pair data in a non-inferiority design). For each reader, both summation and AIP images were not inferior to PR images in terms of being rated identical to PR ($p < 0.05$). Although summation and AIP techniques produce images that differ from PR images, these differences are not easily perceived by radiologists. Thus, the summation or AIP techniques can substitute for PR for the primary interpretation of abdominal CT.

KEY WORDS: Tomography, spiral computed-image processing, computer-assisted-imaging, three-dimensional-image interpretation, computer-assisted-information storage and retrieval

INTRODUCTION

The reduction in the slice thickness of multi-detector-row computed tomography (CT) images has raised several technical challenges such as the data overload¹⁻⁴ and the higher

radiation dose.^{5,6} The conversion of remaining older scanners to 16- or higher detector-row CT will increase such concerns. In view of these two challenges, the summation (averaging multiple contiguous sections) and axial slab average intensity projection (AIP) of thin-section datasets are particularly interesting, as they reduce the number of images and improve image quality without increase in patient dose by averaging pixel values and therefore canceling image noise across the images.⁶⁻⁹ To deal with image overload, many radiologists today are acquiring images using thin

¹From the Department of Radiology, Seoul National University Bundang Hospital, 300 Gumi-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, 463-707, South Korea.

²From the Seoul National University College of Medicine, Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, South Korea.

³From the Division of Multimedia Engineering, College of Information and Media, Seoul Women's University, 126 Gongreung-dong, Nowon-gu, Seoul, 139-774, South Korea.

⁴From the Medical Research Collaborating Center, Seoul National University Hospital, 28 Yongon-dong, Chongno-gu, Seoul, 110-744, South Korea.

⁵From the Seoul National University College of Medicine, Seoul, South Korea.

Correspondence to: Helen Hong, Division of Multimedia Engineering, College of Information and Media, Seoul Women's University, 126 Gongreung-dong, Nowon-gu, Seoul, 139-774, South Korea; tel: +82-2-9705756; fax: +82-31-7874011; e-mail: khlee@radiol.snu.ac.kr

Copyright © 2007 by Society for Imaging Informatics in Medicine

Online publication 6 September 2007

doi: 10.1007/s10278-007-9067-y

collimation and then archiving, filming, and even interpreting images using much thicker (ca 5 mm) section series which are derived from original thin-section datasets by using summation or AIP.⁶⁻⁹ Sliding slab AIP is already widely used for the volumetric navigation of thin-section datasets, in which overlapping slabs of a desired thickness are rapidly rendered creating the illusion of continuity from image to image as the user slides the slab along a viewing direction.^{6,10,11} Regardless of whether summation and AIP are applied at the time of display in an interactive manner, or at the time of archiving or filming, these techniques are being increasingly used,^{6,8,10,12-14} and they are likely to become a primary interpretation mode for thin-section CT datasets, to deal with data overload and to increase image quality.¹⁰

However, because primary reconstruction (PR) directly from raw projection data has been the standard for primary interpretation in clinical practice since the period of single detector-row CT, the likelihood of the utilization of AIP and summation raises a fundamental question, namely, is it justifiable to use summation or AIP for primary interpretation instead of PR? If this is not the case, we have to routinely reconstruct and archive both thick-section (with lower-noise) and thin-section datasets (with higher-resolution) from the same raw projection data for every examination, to take full advantage of modern thin-section scanners.

In this study, we hypothesized that the summation of thin-section image datasets or slab AIP in the axial plane can substitute for thick-section PR images for abdominal applications. To validate our hypothesis, we measured difference between PR images and summation or AIP images and performed a visual image analysis to determine if radiologists can distinguish between the PR images and the summation or AIP images.

MATERIALS AND METHODS

Dataset Preparation for Quantitative Analysis

Our institutional review board approved this study and waived patient consent for the use of clinical images in this study, as the patients' confidentiality was guaranteed by using anonymized images.

This study included 150 consecutive patients (4 to 85 years old, 75 males and 75 females) who underwent single-phase contrast-enhanced abdominal CT in Seoul National University Bundang Hospital during a period of 102 days between April and July 2004. We did not confirm the reasons for CT examination in these patients. Raw projection data were obtained by using 16-detector-row helical CT scanners (Brilliance; Philips Medical Systems, Cleveland, OH, USA) using the following scanning parameters: scan range, from diaphragm to symphysis pubis; detector collimation, 1.5 mm; gantry rotation time, 0.5 s; tube potential, 120 kVp; and pitch, 1.17 to 1.25. Tube current was automatically modulated according to patient body size and the asymmetric nature of the scanned object to give same level of image noise as a reference image (Dose-Right, Philips Medical Systems).

From each raw projection data, two CT image datasets (one thick- and another thin-section image datasets) were reconstructed (Fig. 1). The thick-section image dataset had a section thickness of 5 mm and a reconstruction interval of 4 mm (PR dataset), whereas the thin-section image dataset had a section thickness of 2 mm and a reconstruction interval of 1 mm. All other reconstruction parameters, i.e., image position along the *x*- and *y*-axes, field-of-view (185 to 369 mm), and reconstruction filter type (filter C), were constant for these two image datasets. The reconstruction range along the *z*-axis (297 to 549 mm) was set to be the same for the two image datasets, by trimming both extremes of the scanned range.

From each thin-section image dataset, a summation image dataset (5-mm section thickness, 4-mm reconstruction interval) was calculated by averaging every four images with equal weights, using the CT scanner workstation software (Extended Brilliance Workspace; Philips) (Fig. 2).

For axial slab AIP image datasets, we decided not to use the CT scanner workstation software because the generated AIP images did not match the PR images accurately with respect to image positions along the *z*-axis. In the CT workstation, the desired location of the reformatted image is selected only by moving the cursor in the graphical user interface, and not by specifying an exact location as a numeric value. Therefore, we calculated axial slab AIP image dataset (5-mm

Fig 1. The two CT image datasets reconstructed from a raw projection data. The thick-section image dataset [primary reconstruction (PR) dataset] has a section thickness of 5 mm and a reconstruction interval of 4 mm, whereas the thin-section image dataset has a section thickness of 2 mm and a reconstruction interval of 1 mm. Each *horizontal line* represents an image centered at a given position along the z -axis (e.g., z').

section thickness, 4-mm reconstruction interval) from the thin-section dataset through two steps, resampling and averaging (Fig. 3). In the first step, intermediate images were resampled from thin-section images within a 5-mm-thick slab centered at a given image position along the z -axis. A sample location in an intermediate image was assigned a value calculated by trilinear interpolating¹⁵ the eight voxels (in the thin-section source images) closest to that sample location. The interval along the z -axis between adjacent two intermediate images was equal to the x - (or y -) pixel spacing of the thin-section source images, so that the resampled intermediate images could have isotropic voxels. Finally, an AIP image was obtained by averaging these intermediate images.

To the best of our knowledge, this AIP algorithm is similar to those implemented in most CT scanner workstation softwares.¹⁶

Therefore, theoretical image positions and the nominal section thickness (5-mm) of the summation and AIP image datasets were identical to those of the PR image datasets for each patient. Each type of image dataset contained 74 to 137 (mean $\pm$ SD, 104.5 ± 11.6) images for each patient, totaling 15,677 images for the 150 patients (Fig. 4).

Quantitative Analysis

To analyze differences between PR and summation images, the root-mean-square error (RMSE) was measured as pixel wise difference

Fig 2. Calculation of summation (SUM) image dataset. The SUM image dataset (5-mm section thickness, 4-mm reconstruction interval) was obtained by averaging every four thin-section images. The theoretical image positions and the nominal section thickness (5-mm) of the summation image dataset are identical to those of the primary reconstruction (PR) image dataset in Fig. 1. Each *horizontal line* represents an image centered at a given position along the z -axis (e.g., z').

Fig 3. Calculation of axial slab average intensity projection (AIP) image datasets. The intermediate images (dotted lines) were resampled from thin-section images by trilinear interpolation technique, within a 5-mm-thick slab centered at a given image position along the z-axis (e.g., z'). The final AIP image (5-mm section thickness, 4-mm reconstruction interval) was obtained by averaging these intermediate images. The theoretical image positions and the nominal section thickness (5-mm) of the AIP image dataset are identical to those of the primary reconstruction (PR) image dataset in Fig. 1. Each horizontal line represents an image centered at a given position along the z-axis (e.g., z').

for each of the 15,677 pairs of PR and summation images, using the following equation.

$$\text{RMSE} = \sqrt{\frac{1}{512 \times 512} \sum_{x=0}^{511} \sum_{y=0}^{511} (f(x,y) - g(x,y))^2},$$

where $f(x,y)$ and $g(x,y)$ are the attenuation numbers of x -th row and y -th column pixels in the PR and summation images, respectively. Similarly, the RMSE between each of 15,677 pairs of PR and AIP images was also measured.

Dataset Preparation for Visual Analysis

A visual analysis was performed to determine if radiologists could distinguish between PR and replicate (summation or AIP) images. By generat-

ing a random sequence for the 15,677 PR images, 0–10 (mean $\pm$ SD, 3.33 ± 2.03) PR images per patient were selected to form a 500-image set, and this was subsequently compared with corresponding summation images. Similarly, another set of 500 PR images (0–9 per patient, 3.33 ± 1.92) was selected for comparison with corresponding AIP images. From these two 500-PR-image sets, four groups of 500 image pairs were prepared for visual analysis as follows: group I, the first set of PR images and identical PR images; group II, PR images and corresponding summation images; group III, the second set of PR images and identical PR images; group IV, PR images and corresponding AIP images. Groups I and II were used to compare PR and summation and groups III and IV to compare PR and AIP. Groups I and III served as controls. For groups II

Fig 4. Dataset preparation for quantitative and visual analysis. AIP Axial slab average intensity projection, PR primary reconstruction, SUM summation.

and IV, difference image datasets were also prepared by subtracting replicates from corresponding PR images (Fig. 4).

Visual Analysis

Four board-certified body radiologists, who had 4 to 10 (mean, 7) years of working experience in interpreting abdominal CT findings, participated in this analysis. Each reader was informed of the purpose of the evaluation, a description of the protocol, and the structure of analyzed datasets (Fig. 4).

The 2,000 image pairs of the four groups were rearranged in a random number sequence. Image pairs were presented side-by-side in this order for the two readers, and in the opposite order for the other two readers. Each reader evaluated the 2,000 image pairs in 10 sessions that were separated by at least 7 days.

PR images, which were identified as such, were always placed on the left-hand side, and the test image (replicate image or PR image as control) was displayed on the right. Each reader was asked to independently indicate if the test image on the right was identical to the PR image on the left, or if a detectable difference was present (binary response). This method has been advocated to be highly sensitive by Slone et al.¹⁷ for the evaluation of irreversibly compressed radiographic images, as comparison of a test image with a known original maximizes reader's sensitivity to the difference.¹⁸

During the visual analysis, images were displayed using a standard diagnostic review workstation (DS3000, Impax version 4.5; Agfa HealthCare, Mortsel, Belgium), dual flat-panel monochrome monitors (ME315; Totoku, Tokyo, Japan) with a matrix size of 1,536×2,048 and a diagonal display size of 20.8 in. (52.8 cm), and a matching video hardware (LV32P1; Totoku). Difference images were not presented during the visual analysis. All annotations and labels suggesting image acquisition or postprocessing methods were toggled off. Readers were allowed to magnify images or to adjust window centers and level settings for the image analysis. The ambient room light was subdued. It was recommended for each reader to spend at least 30 s studying each pair of images, but reviewing was conducted at the reader's convenience, without time constraint. A

research assistant supervised each reading session and manually recorded the amount of time needed to read each image pair. Reading time included the time required to move to the next image pair by scrolling and to make a response.

Statistical Analysis

A biostatistician (S.H.) participated in the study design and in the statistical analysis. All analyses were performed using SAS statistical software (version 8.02; SAS Institute, Cary, NC, USA).

Because the distributions of RMSE data were asymmetric, results were reported as medians and percentiles. To compare RMSE between summation and AIP images, the nonparametric Wilcoxon matched-pairs signed-ranks test was used by adjusting for a possible clustering effect in the same patient.¹⁹ The clustering effect (correlation due to clustering) was measured by calculating intraclass correlation coefficient.

To measure a possible clustering effect in the visual analysis results, the intraclass correlation coefficient was calculated for each reader and each group. A Wald-type test statistics for clustered matched-pair data in a non-inferiority design²⁰ was calculated for each of the four dataset groups for each reader. For each reader's results, we tested the hypothesis of inferiority in terms of the probability that test images would be rated to be identical to PR images. The calculated statistic was used to compare groups I and II (for PR vs summation), and groups III and IV (for PR vs AIP). A one-tail *p*-value was used at the significance level of 0.05. This test was used to show that identicalness of a replicate to PR (i.e., the proportion of replicate images rated as being identical to PR images) is no less than some clinically acceptable amount from the assessed identicalness of PR to itself, and therefore, to determine whether the replicate can be accepted as an alternative to PR, considering other merits of the replicate. Contrary to conventional hypothesis tests providing the strength of evidence against the null hypothesis of no 'clinically relevant' difference, the non-inferiority test deals with a null hypothesis stating that the true difference between two procedures is greater than or equal to a prespecified 'clinically unacceptable' threshold. Therefore, when the null hypothesis is rejected, it can be concluded that although a new procedure

might not be as perfect as a standard procedure, but the true difference is still within a clinically acceptable boundary. In this study, we set this clinically unacceptable threshold as -0.01 , on the assumption that radiologists would begin to use summation or AIP if there was no more than a 1% difference between PR images and the replicates, in terms of being rated as being identical to PR images.

RESULTS

RMSE between PR and summation (median = 9.55; 5%, 95% percentiles = 6.68, 13.59) was significantly greater than that between PR and AIP (7.12; 4.26, 11.28) ($p < 0.0001$, intraclass correlation coefficient = 0.61) (Fig. 5).

For visual analysis, the four readers spent 35.9 ± 7.2 (mean $\pm$ SD), 24.3 ± 9.8 , 27.5 ± 10.5 , and 30.1 ± 7.7 s, respectively, for each pair of images in group I; 35.9 ± 6.9 , 23.7 ± 9.7 , 27.4 ± 10.5 , and 29.8 ± 7.4 s for group II; 35.8 ± 7.3 , 23.1 ± 9.5 , 27.9 ± 10.8 , and 30.2 ± 8.1 s for group III; and 35.9 ± 7.4 , 23.4 ± 10.4 , 27.2 ± 9.7 , and 30.4 ± 7.9 s for group IV. The readers' responses for test images rated as being identical to the corresponding PR images are tabulated in Tables 1 and 2. The calculated intraclass correlation coefficients were varied and were as large as 0.54 (in reader 4's responses for group II). For each reader, the non-inferiority test rejected the null hypothesis at 5% significance level and concluded that the summation and AIP images were not more than 1% less rated as being identical to the PR images than the PR images were; the summation and AIP were non-inferior to

Fig 5. Box and whisker plot of the measured RMSEs of 15,677 pairs of primary reconstruction (PR) and summation or axial slab average intensity projection (AIP) images. The middle lines of the boxes show medians, and upper and lower margins of the boxes show upper and lower quartiles, respectively. The ends of the vertical lines show 5 and 95 percentiles.

Table 1. Readers' Visual Analysis Responses

		Reader 1			Reader 2			Reader 3			Reader 4		
		Group II			Group II			Group II			Group II		
Group I	I	I	D	$Z_{EV} = 6.38$	I	D	$Z_{EV} = 6.19$	I	D	$Z_{EV} = 5.73$	I	D	$Z_{EV} = 2.38$
		438	62	$P<0.0001$	498	2	$P<0.0001$	387	93	$P<0.0001$	497	2	$P = 0.008$
	D	0	0		0	0		13	7		1	0	

Data represent numbers of image pairs. Each group includes 500 image pairs as follows. Group I, primary reconstruction (PR) images and identical PR images; group II, PR images and corresponding summation images; group III, PR images and identical PR images; group IV, PR images and corresponding axial slab average intensity projection images. Groups I and III served as controls. *I* Test image was rated as identical to the PR image, *D* test image was rated as being different from the PR image, Z_{EV} Wald-type test statistic for clustered matched-pair data in a non-inferiority design.

the PR in terms of being rated identical to PR (Figs. 6, 7 and 8).

DISCUSSION

In our results, although summation and AIP produced images that differed from PR, they did not contain notable “postprocessing” features that allowed radiologists to distinguish them easily from PR images. These results might provide a basis for the use of summation and AIP as primary archiving or display modes as alternatives to PR for abdominal examinations.¹⁴

Another implication of our results is the justification of using interactive, sliding slab AIP technique for primary interpretation. Sliding slab mode operates by rendering, within the acquired volume data, only those portions that lie between a set of parallel clipping planes oriented perpendicular to the chosen viewing direction. This is a fast and efficient technique for reviewing a large thin-section dataset, as overlapping slabs are rendered

rapidly, creating the illusion of continuity from image to image as the user slides the slab along the viewing direction.^{10,11} If average intensity projection is used for the sliding slab, image quality can be also improved by averaging a series of adjacent data points perpendicular to the reformatted plane, while keeping the patient dose low (even if the source images are acquired in thin-section)^{10,13,21} and preserving the spatial resolution inherent in the source transverse images.^{10,21,22} If something suspicious is detected, thinner source images can be examined easily for a more detailed view.^{10,21}

If we do not have to always reconstruct, archive, and review a thick-section PR dataset in addition to a thin-section dataset from the same raw projection data, productivity would be enhanced, and radiologists would be able to concentrate on navigating a single volumetric dataset by using interactive sliding slab viewing mode. If current and previous images (assuming that previous images are electronically stored) could be compared on a routine basis in interactive sliding slab viewing mode, productivity would be further

Table 2. Readers' Visual Analysis Responses

		Reader 1			Reader 2			Reader 3			Reader 4		
		Group IV			Group IV			Group IV			Group IV		
Group III	I	I	D	$Z_{EV} = 5.64$	I	D	$Z_{EV} = 2.82$	I	D	$Z_{EV} = 6.23$	I	D	$Z_{EV} = 1.86$
		459	41	$P<0.0001$	498	1	$P = 0.002$	409	76	$P<0.0001$	494	3	$P = 0.03$
	D	0	0		1	0		10	5		3	0	

Data represent numbers of image pairs. Each group includes 500 image pairs as follows. Group I, primary reconstruction (PR) images and identical PR images; group II, PR images and corresponding summation images; group III, primary reconstruction (PR) images and identical PR images; group IV, PR images and corresponding axial slab average intensity projection images. Groups I and III served as controls. *I* Test image was rated as identical to the PR image, *D* test image was rated as being different from the PR image, Z_{EV} Wald-type test statistic for clustered matched-pair data in a non-inferiority design.

Fig 6. Primary reconstruction (PR), summation, axial slab average intensity projection (AIP), and their difference images in a 46-year-old male. **a** PR. **b** Summation. **c** AIP. **d** Difference image of PR vs summation. **e** Difference image of PR vs AIP. The summation and AIP images are not easily discerned from the PR image. All readers rated the summation image as being identical to the PR image. Window width and level are 435 and 15 HU for (a), (b), and (c).

Fig 7. Region-of-interest in primary reconstruction (PR), summation, axial slab average intensity projection (AIP), and their difference images in a 50-year-old female. **a** PR. **b** Summation. **c** AIP. **d** Difference image of PR vs summation. **e** Difference image of PR vs AIP. Although the diaphragm (*arrowheads*) and the subphrenic fat (*arrow*) are more prominent in the summation and AIP images than in the PR images, the summation and AIP images are not easily discerned from the PR image. Two readers (readers 1 and 3) rated the summation image as being different from the PR image. Window width and level are 435 and 15 HU for (a), (b), and (c).

Fig 8. Region-of-interest in primary reconstruction (PR), summation, axial slab average intensity projection (AIP), and their difference images in a 35-year-old male. **a** PR. **b** Summation. **c** AIP. **d** Difference image of PR vs summation. **e** Difference image of PR vs AIP. Although the folds in the small bowel are more prominent in the summation and AIP images than in the PR images (arrows), the summation and AIP images are not easily discerned from the PR image. Two readers (readers 1 and 3) rated the AIP image as being different from the PR image. Window width and level are 435 and 15 HU for (a), (b), and (c).

increased. We believe that radiologists will accept such an image reviewing system as reasonable and begin to use it for primary interpretation.

It is not surprising that summation and AIP produced mathematically different images from thick-section PR (Figs. 6, 7 and 8). Although their nominal section thicknesses are the same (5-mm), the shapes of their section-sensitivity profiles (how much a point in the object contributes to the image as a function of its distance from the center of the section) are different because the section-sensitivity profiles of summation and AIP images are affected by averaging and resampling processes along the z-axis between contiguous thin-section images, whereas the section-sensitivity profile of PR images is determined by the image reconstruction algorithm and filter.

Our quantitative analysis showed that the RMSE between PR and summation was significantly greater than that between PR and AIP. Therefore, the image fidelity of AIP is higher than that of summation, if PR is regarded as the standard of practice.

We limited this study to axial plane because it is impossible to generate PR images in non-axial plane which can be compared with summation or AIP images. Although commercial scanners can now generate coronal or sagittal images, these are not PR but AIP from the transverse thin-section PR datasets. We believe our results can provide a basis for justification of using such slab multi-planar reformations in clinical practice, as the principle of AIP is the same regardless of imaging plane.

It is worthwhile to mention several considerations regarding our study sample. First, with an intention to generalize our results throughout the abdomen, we compared the images of 150 patients who underwent abdominal CT for diverse reasons. Consequently, many images necessarily contained only normal structures. One might argue that our results cannot be generalized to cover a broad range of potential abnormalities in abdominal CT and that many interpretative or diagnostic studies might be required to accept our hypothesis that the summation or AIP techniques can adequately

substitute for PR. However, we chose a visually based image comparison study rather than an interpretative or diagnostic study of a single disease because we believe that the former approach would be more conservative in terms of accepting our hypothesis. It is known that radiologists are able to make accurate diagnoses from images containing minor artifacts while they are extremely sensitive to visual differences between images.²³ It is very unlikely that the visual differences are pronounced particularly in the areas of abnormalities. Therefore, we believe that our results would be reproducible even with a study sample containing more abnormalities.

If we had included all images in a smaller number of patients in the visual analysis, we could have avoided missing any critical images by random image sampling, without increase in the sample size. However, reducing the number of patients might have increased the clustering effect in the statistical analysis and decreased the generalizability of our results regarding individual variation between patients.

Second, the sample size for the visual analysis could have been significantly reduced if we had used corresponding images of summation, AIP, and PR. However, such study design would have complicated the statistical analysis. As we are not aware of any feasible statistical method to test non-inferiority for clustered data between three groups, we chose a more straightforward study design—separating groups I–II comparison and groups III–IV comparison.

Third, we limited our visual analysis to the selected 500 images because of our limited research resources, while we performed the quantitative analysis in 15,677 images. Although a statistical significance might have been obtained even with a smaller sample, there was no need to discard a majority of the dataset in the quantitative analysis that could be performed much easier than the human visual analysis.

The limitations of the present study are as follows. First, tested images were reconstructed with fixed reconstruction parameters using a single scanner type. To further generalize our results, more studies are needed to determine the effects of varying reconstruction parameters and scanner types. Mathematical simulations and phantom studies might be also helpful in terms of generalizing. Second, the number of readers was too small

to allow generalizing our results to the population of readers. In view of the variability in the response pattern shown by the four readers (Tables 1 and 2), a larger study might be needed to verify our results. Third, we did not analyze any pattern(s) of differences between PR and summation or AIP because the purpose of this study was to assess the magnitude of the differences. Fourth, some under- or overestimation of the statistical significances might occur in the analysis method²⁰ which we adopted in the visual analysis. When a reader's positive response (the two compared images are different) was rare, most of the clusters could return a result of 'no difference' rather than 'difference' between two groups, and therefore, averaging these results at the cluster level might exaggerate the statistical significance by producing a very small *p*-value. When the positive response was extremely rare so that few clusters could contain the positive response, the clustering effect could be inflated, which might then rather magnify the *p*-value. We postulate that these artifacts can explain some discrepancy between the *p*-values and the apparent statistical significances that can be expected from the data tabulated in Tables 1 and 2. Nevertheless, we believe that it was still reasonable to employ an analysis method taking into consideration the clustering effect and that our results provide considerable evidence to support the non-inferiority.

CONCLUSION

Although summation and axial slab AIP techniques of thin-section abdominal CT datasets produce images that differ from thick-section PR images, these differences are not easily perceived by radiologists. On the assumption that radiologists are able to make accurate diagnoses from images with these imperceptible differences, the summation or AIP techniques can be substituted for PR for the primary interpretation of abdominal CT. We believe our results can also provide a basis for justification of using slab multiplanar reformations in clinical practice.

ACKNOWLEDGMENT

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-

2006-311-D00168). We thank the radiologists who participated as readers and Sang Hyun Kim, R.T. for his assistance during image dataset preparation.

REFERENCES

1. Rubin GD: Data explosion: the challenge of multidetector-row CT. *Eur J Radiol* 36:74–80, 2000
2. Rubin GD: 3-D imaging with MDCT. *Eur J Radiol* 45 (Suppl 1):S37–S41, 2003
3. Tamm EP, Thompson S, Venable SL, McEnery K: Impact of multislice CT on PACS resources. *J Digit Imaging* 15 Suppl 1:96–101, 2002
4. Lee KH, Lee HJ, Kim JH, Kang HS, Lee KW, Hong H, Chin HJ, Ha KS: Managing the CT data explosion: initial experiences of archiving volumetric datasets in a mini-PACS. *J Digit Imaging* 18:188–195, 2005
5. McNitt-Gray MF: AAPM/RSNA Physics tutorial for residents: topics in CT. Radiation dose in CT. *Radiographics* 22:1541–1553, 2002
6. Dalrymple NC, Prasad SR, Freckleton MW, Chintapalli KN: Informatics in radiology (infoRAD): introduction to the language of three-dimensional imaging with multidetector CT. *Radiographics* 25:1409–1428, 2005
7. Ba-Ssalamah A, Prokop M, Uffmann M, Pokieser P, Teleky B, Lechner G: Dedicated multidetector CT of the stomach: spectrum of diseases. *Radiographics* 23:625–644, 2003
8. Prokop M: Spiral and multislice computed tomography of the body, Stuttgart, Germany: Georg Tieme Verlag, 2003
9. Siegel E: SCAR University syllabus at the 2004 annual meeting, Great Falls, VA: Society for Computer Applications in Radiology, 2004
10. Lee KH, Kim YH, Hahn S, Lee KW, Kim TJ, Kang S-B, Shin JH: CT diagnosis of acute appendicitis: advantages of reviewing thin-section datasets using sliding slab average intensity projection technique. *Invest Radiol* 41:579–585, 2006
11. Napel S, Rubin GD, Jeffrey RB, Jr.: STS-MIP: a new reconstruction technique for CT of the chest. *J Comput Assist Tomogr* 17:832–838, 1993
12. van Ooijen PM, Ho KY, Dorgelo J, Oudkerk M: Coronary artery imaging with multidetector CT: visualization issues. *Radiographics* 23:e16, 2003
13. Prokop M: Multislice CT: technical principles and future trends. *Eur Radiol* 13(Suppl 5):M3–M13, 2003
14. Jeong DK, Lee KH, Kim BH, Kim KJ, Kim YH, Bajpai V, Shin YG: On-the-fly generation of multiplanar reformation images independent of CT scanner type. *J Digit Imaging* (in press), 2007 DOI [10.1007/s10278-007-9032-9](https://doi.org/10.1007/s10278-007-9032-9)
15. Wolberg G: Digital Image Warping. Los Alamitos, CA: IEEE Computer Society Press, 1990
16. Venema HW, Phoa SS, Mirck PG, Hulsmans FJ, Majoie CB, Verbeeten B, Jr.: Petrosal bone: coronal reconstructions from axial spiral CT data obtained with 0.5-mm collimation can replace direct coronal sequential CT scans. *Radiology* 213:375–382, 1999
17. Slone RM, Foos DH, Whiting BR, Muka E, Rubin DA, Pilgram TK, Kohm KS, Young SS, Ho P, Hendrickson DD: Assessment of visually lossless irreversible image compression: comparison of three methods by using an image-comparison workstation. *Radiology* 215:543–553, 2000
18. Gur D, Rubin DA, Kart BH, Peterson AM, Fuhrman CR, Rockette HE, King JL: Forced choice and ordinal discrete rating assessment of image quality: a comparison. *J Digit Imaging* 10:103–107, 1997
19. Larocque D: Statistical modeling and analysis for complex data problem. Berlin Heidelberg New York: Springer, 2005
20. Durkalski VL, Palesch YY, Lipsitz SR, Rust PF: Analysis of clustered matched-pair data for a non-inferiority study design. *Stat Med* 22:279–290, 2003
21. Cody DD: AAPM/RSNA physics tutorial for residents: topics in CT. Image processing in CT. *Radiographics* 22:1255–1268, 2002
22. Jeong YJ, Lee KS, Yoon YC, Kim TS, Chung MJ, Kim S: Evaluation of small pulmonary arteries by 16-slice multidetector computed tomography: optimum slab thickness in condensing transaxial images converted into maximum intensity projection images. *J Comput Assist Tomogr* 28:195–203, 2004
23. Slone RM, Muka E, Pilgram TK: Irreversible JPEG compression of digital chest radiographs for primary interpretation: assessment of visually lossless threshold. *Radiology* 228:425–429, 2003