
Improved Prognostic Value of ^{18}F -FDG PET Using a Simple Visual Analysis of Tumor Characteristics in Patients with Cervical Cancer

Tom R. Miller, MD, PhD¹; Edward Pinkus, MD¹; Farrokh Dehdashti, MD¹; and Perry W. Grigsby, MD²

¹Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, Missouri; and ²Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri

In patients with cervical cancer, it is important to estimate prognosis at the time of diagnosis. This study using PET with ^{18}F -FDG was undertaken to determine whether a simple and fast visual analysis of characteristics of the primary tumor before initiation of treatment could achieve this goal. **Methods:** Forty-seven patients with cervical cancer who were to be treated by combined radiation therapy and chemotherapy were imaged before beginning treatment. They were then followed for up to 3 y for evidence of recurrence or death. Images of the chest, abdomen, and pelvis were obtained 40–90 min after administration of 370–555 MBq (10–15 mCi) ^{18}F -FDG. Three observers then independently graded the primary tumor for size (0 = small, 1 = moderate, 2 = large), shape (0 = spherical, 1 = nonspherical), heterogeneity of uptake (0 = none, 1 = moderate, 2 = marked), and presence of lymph nodes (0 = none, 1 = pelvic, 2 = paraaortic, 3 = distant). The scores were summed to achieve a total score. A statistical calculation demonstrated that a score cutoff of 4 best separated patients with a good prognosis from patients with a bad prognosis. Kaplan–Meier analysis was used to compute progression-free survival and overall survival. Evaluation of lymph nodes alone was compared with the grading of tumor characteristics. **Results:** Observers 1 and 2 scored 26 patients as having a good prognosis and 21 as having a bad prognosis. Observer 3 scored 30 and 17, respectively, a statistically insignificant difference. Survival curves were almost identical for the 3 observers. For progression-free survival, approximately 12% of patients with a good score had disease recurrence whereas approximately 75% with a bad score had disease recurrence. For overall survival, approximately 10% (good) and 80% (bad) died. Evaluation of lymph nodes also separated the groups, but not as well as did visual analysis alone. The combination of the 2 was only slightly superior to visual assessment alone. **Conclusion:** A simple, rapid, and highly reproducible system is described for visual grading of characteristics of the primary tumor in patients with cervical cancer at the time of diagnosis. This approach separates pa-

tients with a poor prognosis from those who will do well, thus providing a new tool for accurate estimation of prognosis.

Key Words: PET; cervical cancer; ^{18}F -FDG

J Nucl Med 2003; 44:192–197

Treatment planning and determination of prognosis are important parts of the initial evaluation of patients with cancer. In cervical cancer, these goals have traditionally been achieved by clinical assessment based on the International Federation of Gynecology and Obstetrics (FIGO) staging system (1). In recent years, modern imaging techniques, including CT and MRI, have contributed to improved assessment of these patients (2–5). Reports from our institution (6,7) and elsewhere (8–11) recently demonstrated substantial value for PET using the radiopharmaceutical ^{18}F -FDG. The primary tumor is well seen, and lymph nodes are reliably evaluated with greater accuracy than by CT, MRI, or lymphangiography (6,7,11). We demonstrated further improvement in assessment of prognosis by ^{18}F -FDG PET through addition of quantitative determination of the volume of the primary tumor (12). Unfortunately, this quantitative measurement is somewhat time consuming and requires specialized software. Thus, we undertook this study using only visual, qualitative evaluation of ^{18}F -FDG PET images to determine whether a simple, fast, and reproducible method could be developed that would more effectively direct therapy and estimate prognosis in patients with cervical cancer.

MATERIALS AND METHODS

Forty-seven consecutive patients with cervical cancer diagnosed between January 1998 and September 1999 whose image data were available were retrospectively evaluated. They all had biopsy-proven cervical cancer and were to undergo primary treatment by radiation therapy and, in 28 patients, with adjuvant cisplatin chemotherapy (13). The radiation therapy consisted of external-beam irradiation over 6 wk with 2 intracavitary (brachytherapy) treatments during the external-beam therapy. The total dose to point A was 85 Gy, and the maximum doses to the bladder,

Received Apr. 24, 2002; revision accepted Sep. 17, 2002.
For correspondence contact: Tom R. Miller, MD, PhD, Division of Nuclear Medicine, Mallinckrodt Institute of Radiology, 510 S. Kingshighway Blvd., St. Louis, MO 63110.
E-mail: millert@mir.wustl.edu

rectum, and lateral surface of the vagina were 75, 70, and 130 Gy, respectively.

All patients underwent ^{18}F -FDG PET before the beginning of treatment following the standard imaging protocol in our department, consisting of fasting for at least 4 h followed by placement of a Foley catheter, intravenous hydration, and administration of 20 mg of furosemide to minimize bladder activity. Then, 370–555 MBq (10–15 mCi) ^{18}F -FDG were administered, and images of the chest, abdomen, and pelvis were obtained beginning 40–90 min later. A 47-slice ECAT-EXACT PET tomograph (Siemens/CTI, Knoxville, TN) was used, with 10-min emission and 2-min transmission images collected at 4–6 bed positions. A segmentation algorithm was used to generate a transmission map (14) for attenuation correction, and transaxial slices were reconstructed with use of the ordered-subsets expectation maximization algorithm (15) and Butterworth filtering. The data were reconstructed in a 128×128 pixel matrix with a pixel size and slice spacing of 4.3 mm. The reconstructed spatial resolution was typically 8 mm in full width at half maximum.

Three observers with different levels of experience independently evaluated the ^{18}F -FDG PET images. Two were attending nuclear medicine physicians with many years of experience, and the third was a radiologist in nuclear medicine fellowship training. The images were evaluated without knowledge of the clinical history or subsequent follow-up except that all patients had newly diagnosed cervical cancer. The data were interactively viewed on a computer monitor with display of transaxial, coronal, and sagittal slices. Gray scale and various color scales could be selected and adjusted.

Three characteristics of the primary tumor were evaluated and numerically graded: size, shape, and heterogeneity of ^{18}F -FDG uptake (Table 1). Size was scored as 0 for small, 1 for moderate, or 2 for large. Small size was defined as tumor occupying only a small portion of the lower pelvis, with diameter ≤ 4 cm (10 pixels); moderate, as tumor occupying most of the lower pelvis; and large, as tumor outside the lower pelvis, with diameter > 10 cm (25 pixels). Shape was scored as 0 for spherical or 1 for nonspherical, typically tubular. Heterogeneity was scored as 0 for none, 1 for moderate, or 2 for marked, where moderate heterogeneity was approximately a 10%–30% count variation across the tumor and marked was a variation $> 30\%$. Lymph node involvement was also scored, using 0 for none, 1 for pelvic, 2 for paraaortic, and 3 for distant. If nodes were present in more than a single region, the highest score was used. Metastases to other organs were not observed. Thus, scores could range from a minimum of 0 (small, spherical, homogeneous, no lymph node involvement) to a maximum of 8 (large, nonspherical, markedly heterogeneous, distant nodal metastases).

Clinical examinations were performed 6 wk after the end of treatment and thereafter every 3 mo. No patients were lost to follow-up.

Two different survival endpoints were used: progression-free survival (PFS) and overall survival (OS). PFS was the interval from the beginning of radiation therapy to the time of the first recurrence or the last follow-up visit. OS was the time from the first treatment to death or the last follow-up visit. The Kaplan–Meier method was used for survival analysis (16), with statistical significance assessed by the log-rank test.

To determine the best cutoff value for the overall image score, we performed the Kaplan–Meier analysis for several different cutoff values, and the fraction of patients with and without recurrence or death was computed at the end of the follow-up period. The best cutoff was then selected following an approach similar to that used in conventional evaluation and diagnostic tests. The likelihood ratios (sensitivity/[1 – specificity]) for different cutoff values was computed, where sensitivity was the true-positive fraction for a positive cutoff value predicting recurrence or death and specificity was the true-negative fraction (17).

The scores of the 3 readers were compared using a 2×2 contingency table and the McNemar test (18). The relative performance of the observers was also evaluated by inspection of the survival curves generated from their independent interpretations.

Finally, survival was computed with consideration of only lymph node status and with consideration of only visual scoring, without addition of the lymph node information.

RESULTS

The 47 patients were drawn from the group used in our recent tumor-volume article (12) (digital data were no longer available for 4 patients from the previous group). The quantitative results of that work were not known to the observers in this study. The patients ranged from 24 to 84 y old (mean, 48 y), and they were followed for 0.2–3.1 y (mean, 1.7 y). Patient characteristics are shown in Table 2. Typical examples of patient studies are shown in Figure 1.

A score cutoff of 4 best separated the patients with good survival from those with poor survival. Thus, studies were graded as indicating a good prognosis when scores were < 4 and as indicating a bad prognosis when scores were ≥ 4 .

Figure 2 shows the Kaplan–Meier graphs of PFS and OS for observer one. This observer scored 26 patients as having a good prognosis and 21 as having a bad prognosis. The respective numbers were 26 and 21 patients for observer 2 and 30 and 17 patients for observer 3.

TABLE 1
Scoring System

Size	Shape	Heterogeneity	Lymph nodes*
0 = small (< 4 cm)	0 = spherical	0 = none	0 = none
1 = moderate (4–10 cm)	1 = nonspherical	1 = moderate (10%–30%)	1 = pelvic
2 = large (> 10 cm)	—	2 = marked ($> 30\%$)	2 = paraaortic
—	—	—	3 = distant

*Highest value is used as score.

TABLE 2
Patient Characteristics

Patient no.	Stage	Histology	Age (y)	Status	PFS (d)	OS (d)
1	IIb	Squamous	43	AWD	622	832
2	IIIb	Squamous	45	DOD	96	322
3	IIb	Squamous	61	DOD	151	342
4	IIb	Adenosq	44	DOD	95	466
5	IIb	Squamous	27	DOD	211	275
6	Ib2	Adenosq	56	DOD	337	397
7	IIIb	Squamous	52	DOD	103	384
8	Ib2	Squamous	24	DOD	188	270
9	IVb	Squamous	38	DOD	132	149
10	IIIb	Squamous	40	DOD	80	138
11	IIb	Squamous	67	DOD	196	252
12	IIIb	Squamous	47	DOD	55	55
13	IIb	Squamous	43	DOD	160	177
14	IIb	Squamous	52	DOD	213	239
15	Ib2	Squamous	45	DOD	265	730
16	IIIb	Squamous	34	DOD	99	334
17	IIIb	Squamous	48	DOD	345	429
18	IIb	Squamous	47	NED	1,017	1,017
19	Ib2	Squamous	28	NED	498	498
20	IIb	Squamous	28	NED	615	615
21	IIIb	Squamous	31	NED	616	616
22	IIIb	Squamous	47	NED	1,131	1,131
23	IIb	Squamous	54	NED	894	894
24	IIIb	Adenosq	49	NED	704	704
25	IIb	Squamous	58	NED	732	732
26	IIIb	Squamous	27	NED	752	752
27	Ib1	Squamous	47	NED	961	961
28	IIb	Squamous	55	NED	703	703
29	Ia1	Adeno	53	NED	968	968
30	IIb	Squamous	47	NED	643	643
31	IIb	Adeno	45	NED	883	1,040
32	IIb	Squamous	83	NED	661	661
33	IIb	Squamous	42	NED	571	571
34	Ib2	Squamous	47	NED	703	703
35	IIb	Squamous	48	NED	1,015	1,015
36	IIIb	Squamous	71	NED	596	596
37	Ib2	Squamous	29	NED	799	799
38	IIb	Squamous	43	NED	826	826
39	IIIb	Squamous	44	NED	761	761
40	IIb	Squamous	56	NED	830	830
41	IIb	Squamous	57	NED	730	730
42	IIb	Squamous	54	NED	614	614
43	Ib2	Squamous	60	NED	902	902
44	IIb	Squamous	87	NED	154	154
45	IIb	Squamous	63	NED	622	622
46	Ib1	Squamous	71	NED	788	788
47	Ib2	Squamous	36	NED	382	382

AWD = alive with disease; DOD = dead of disease; Adenosq = adenosquamous; NED = no evidence of disease; Adeno = adenocarcinoma.

As an illustration of the computation of likelihood ratios to determine the best cutoff value for scoring, consider the PFS graph shown in Figure 2A, where the cutoff value is 4. At the right end of the graph, the true-positive fraction, or sensitivity (fraction of patients graded as having a bad prognosis and in whom disease recurred), is seen to be 0.69

($1 - \text{survival}$), and the true-negative fraction, or specificity (fraction of good-prognosis patients in whom disease did not recur), is seen to be 0.87. Thus, the likelihood ratio for this patient is 5.3 ($0.69/[1 - 0.87]$).

The complete set of likelihood ratios for observer 1 computed for representative cutoff values is shown in Table 3; the results for the other observers were almost identical. Thus, for both PFS and OS, a total image score of 4 or greater gave the best separation between the patients with and the patients without recurrence or death.

The performance of the 3 observers was compared using contingency tables for each combination of observers, as shown in Table 4. Differences in all the pairs of observers were insignificant.

The reproducibility of the readings of the 3 observers is shown most dramatically in Figure 3, where the survival curves for the 3 observers are superimposed. Note the almost identical results for the observers' independent readings of the studies.

Figure 4 shows the survival graphs using only the lymph node data. The studies were scored as showing the presence or absence of lymph node disease. Although the curves separated the patients who did well from those who did poorly ($P = 0.04$ and $P = 0.03$ for PFS and OS, respectively), the distinctions were much less pronounced than when the tumor-grading scheme was added (Figs. 2 and 3). Separate consideration of lymph node involvement at the paraaortic or distant level appeared to separate the groups better, but only 7 patients had disease at these higher levels. Figure 5 shows the graphs for the visual grading of tumor characteristics when the lymph node score was excluded from the scoring. In that case, a score of 3 best separated the

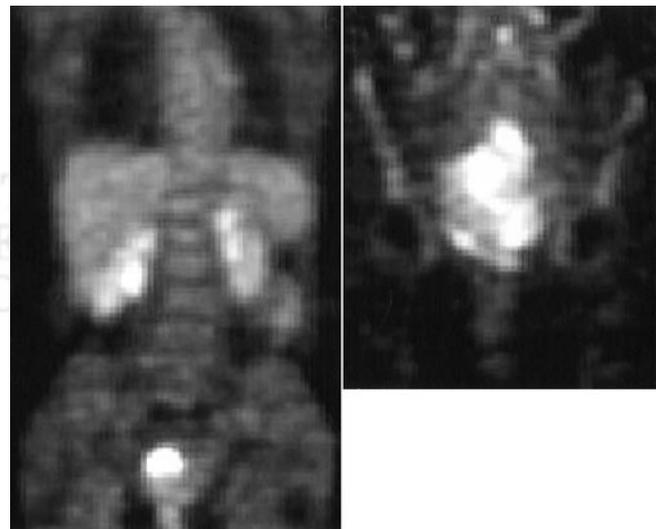


FIGURE 1. On left is coronal image of patient with small, spherical, and homogeneous primary tumor and no lymph node disease (score = 0), who was alive without disease at 703 d. On right is coronal image of patient with large, nonspherical, and markedly heterogeneous primary tumor (score = 2 + 1 + 2 + 2, or 7), who was dead of her disease at 149 d. Patient on right also had paraaortic lymph node disease (not shown).

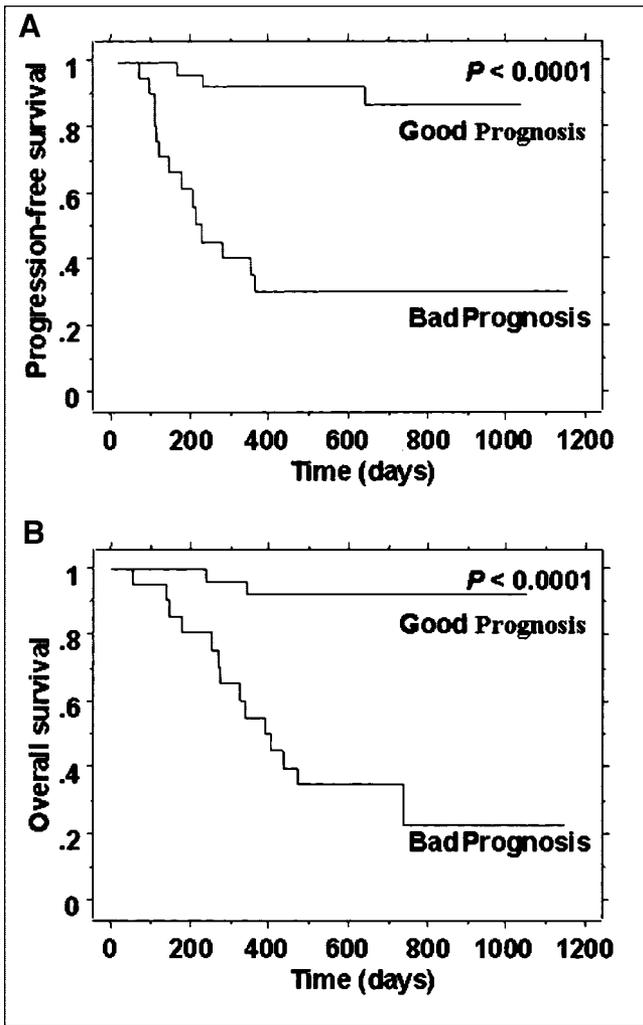


FIGURE 2. (A) PFS for scoring of observer 1. (B) OS for observer 1. Good Prognosis = patients with total score < 4; Bad Prognosis = patients with score ≥ 4.

groups, as determined by the likelihood-ratio methodology. Note that visual analysis alone (Fig. 5) gave results superior to lymph node grading (Fig. 4) and almost as good as the use of both together (Figs. 2 and 3).

DISCUSSION

Cervical cancer is staged by the FIGO system, which is based primarily on physical examination (1). Assessment of prognosis is known to be relatively unreliable by this approach. Thus, imaging methods, including CT, MRI (2-5),

TABLE 3

Likelihood Ratios as Function of Cutoff Value for Observer 1

Cutoff	PFS	OS
3	4.8	9.0
4	5.3	9.5
5	3.9	5.4
6	3.6	4.0

TABLE 4

Comparison of Numbers of Patients Scored as Having Bad Prognosis and Good Prognosis by the 3 Pairs of Observers

Observer pair	Bad prognosis	Good prognosis
2/1		
Bad prognosis	20	1
Good prognosis	1	25
3/1		
Bad prognosis	17	4
Good prognosis	0	26
3/2		
Bad prognosis	17	4
Good prognosis	0	26

and PET (6-11), have been investigated with emphasis on detection of lymph nodes and measurement of the volume of the primary tumor.

In a recent report (12), our group demonstrated that tumor volume could be accurately quantified by ¹⁸F-FDG PET. For

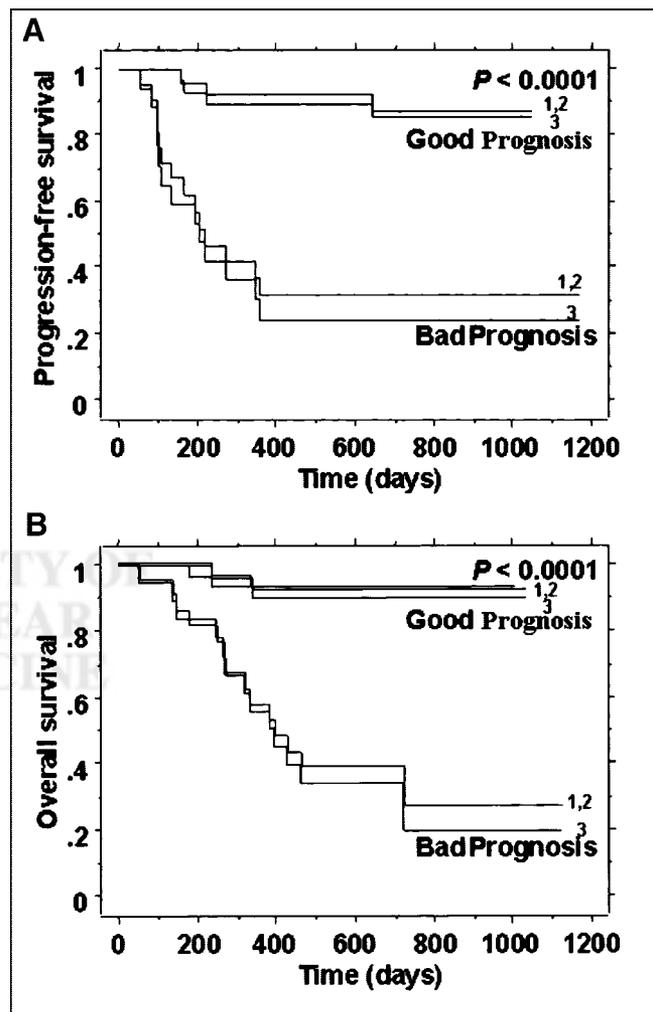


FIGURE 3. (A) PFS superimposed for the 3 observers. (B) OS for the 3 observers. Good Prognosis = patients with total score < 4; Bad Prognosis = patients with score ≥ 4.

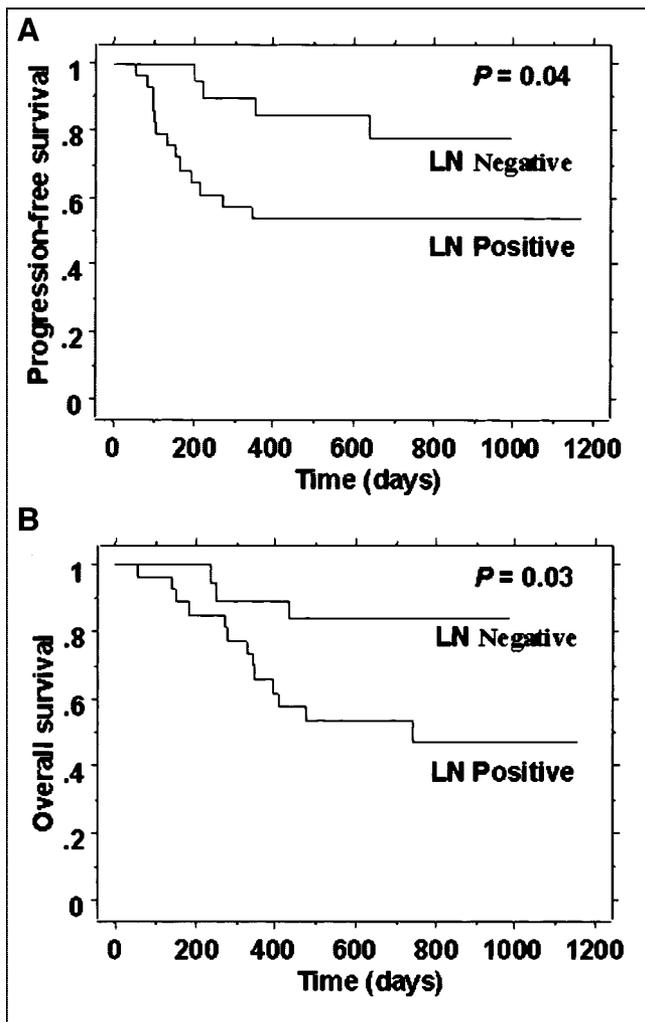


FIGURE 4. (A) PFS when only lymph node status was considered. (B) OS when only lymph node status was considered. LN negative = no lymph node disease; LN positive = disease at any site.

both PFS and OS, patients were well separated according to whether the tumor volume was more or less than 60 cm³. Although this method was effective, the quantification requires special 3-dimensional software not generally available and, in a small fraction of studies, a somewhat cumbersome manual editing process to eliminate intense bladder activity. Thus, here we investigated a completely visual approach to the evaluation of the PET images, requiring no special computer processing and using only the conventional multislice display software available on all PET systems.

This visual analysis of course included the volume—quickly and easily estimated as small, medium, or large. Our experience with many PET studies on cervical cancer has suggested that tumor shape, either spherical or more tubular, is readily evaluated and potentially important. Because cervical tumors vary markedly in the uniformity of ¹⁸F-FDG uptake, an estimate of heterogeneity was included. This nonuniformity is not surprising in view of the potential for

regional tumor necrosis, variability in oxygen saturation, and other metabolic factors.

The most important finding of this work was that this visual grading system predicted progression and OS with high accuracy from only a single PET scan before initiation of treatment (Figs. 2 and 3). Indeed, this approach actually yielded better results than did the much more difficult method previously described (12). For PFS, the good-prognosis and bad-prognosis groups were separated into approximately 10% and 50% recurrence in that report but had approximately 10% and 70% separation in the current report (Fig. 3). For OS, the results of the present study were also superior to those reported previously. In the previous study, FIGO stage was not found to predict recurrence or death.

A second important observation was that this evaluation was performed qualitatively, requiring only an additional 1 or 2 min of time without use of specialized software. Although the description of the scoring system (Table 1)

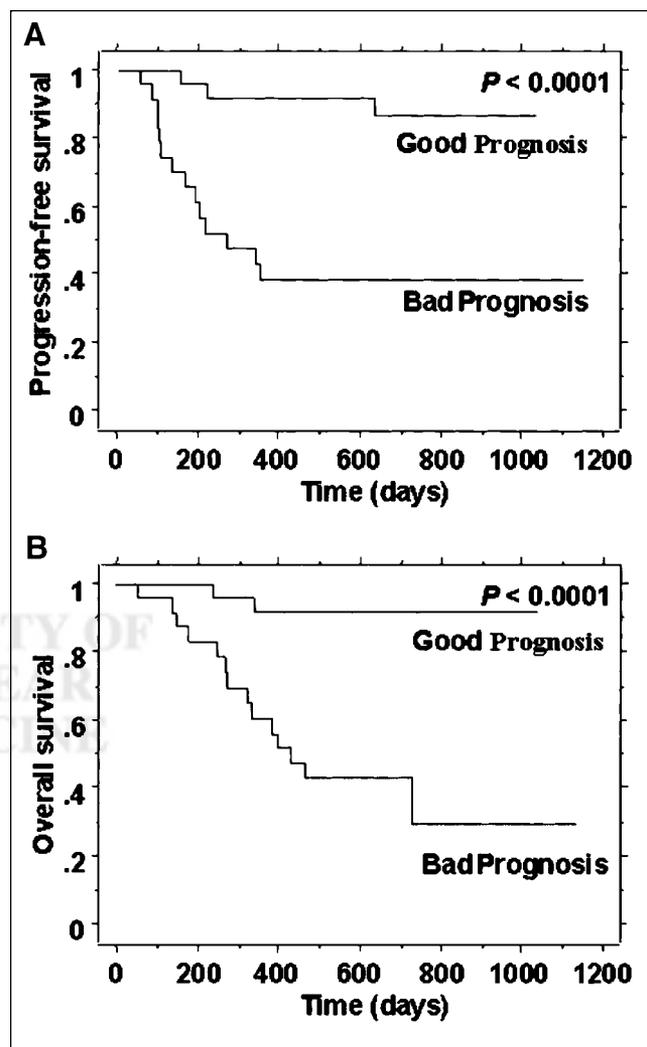


FIGURE 5. (A) PFS for scoring of observer 1 when only visual characteristics of tumor were considered. (B) OS for visual characteristic scoring of observer 1. Good Prognosis = patients with total score < 3; Bad Prognosis = patients with score \geq 3.

appears somewhat quantitative, the grading was, in fact, entirely qualitative and rapid.

Finally, the method was highly reproducible, as shown by the almost identical results achieved by 3 observers having differing levels of experience and interpreting the images independently (Fig. 3; Table 4). Thus, this fast, simple approach could be implemented in any department.

The grading system also incorporated lymph node involvement by tumor, an important prognostic factor documented in a recently reported PET study from our laboratory (7). Although lymph node assessment is a fundamental component in the interpretation of most tumor-imaging studies, we found here that the addition of 3 easily evaluated tumor characteristics significantly improved the estimation of prognosis. A comparison of Figures 4 and 5 shows that the new visual grading system performed better than did analysis of lymph nodes alone. Combining the visual score with lymph node findings (Figs. 2 and 3) led to only a small additional improvement.

The robust nature of this grading scheme appears to arise partly from the correlation of the factors: As tumors grow larger, they tend to become less spherical and more heterogeneous. However, analysis of the factors separately showed a less significant separation of patients and a poorer interobserver correlation than did combined analysis.

A recent study (19) mathematically analyzed the findings described in a large number of articles on measurement or estimation of the tumor volume of many types of cancer by a wide variety of imaging techniques. That study showed that the volume of the tumor is the key factor in prognosis for many cancers. The present work and our previous quantitative article support this conclusion (12). The mathematic analysis of many techniques with quite variable accuracies showed that precision in quantification may not be essential. That observation was also confirmed by this work, in which 3 observers achieved almost identical results from qualitative assessments.

The remarkable quantitative power of PET has led many investigators to emphasize methods based on numeric computation—an approach we followed in our previous study on cervical cancer. The work presented here, using visual assessment by trained observers, lies on a path set out many years ago by two ¹⁸F-FDG PET pioneers (20) and is in the spirit of most of medical imaging, which continues to rely on qualitative judgment. Indeed, this method, involving subjective scoring of multiple factors, is somewhat reminiscent of the general grading approach widely used in medicine, such as the American Pediatric Gross Assessment Record used to evaluate newborn infants (21). The PET technique presented here may have applications in other cancers, perhaps by analysis of the same or different characteristics of the tumor.

CONCLUSION

A system is described for visually scoring characteristics of the primary tumor in patients with cervical cancer. This

rapid and simple approach leads to a highly accurate and reproducible assessment of prognosis from a single PET scan before initiation of treatment. Although assessment of lymph nodes is valuable, this method appears to perform better than lymph node evaluation and is slightly more effective when the two are combined.

ACKNOWLEDGMENT

This project was supported in part by grant R01 CA85797 from the National Institutes of Health. This work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Greene FL, Page DL, Fleming ID, et al. *AJCC Cancer Staging Manual*. New York, NY: Springer; 2002:259–262.
2. Suyama S, Nakaguchi T, Kawakami K, et al. Computed tomography analysis of causes of local failure in radiotherapy for cervical carcinoma. *Cancer*. 1998;83:1956–1965.
3. Mayr N, Yuh W, Zheng J, et al. Tumor size evaluated by pelvic examination compared with 3-D MR quantitative analysis in the prediction of outcome for cervical cancer. *Int J Radiat Oncol Biol Phys*. 1997;39:395–404.
4. Yu KK, Forstner R, Hricak H. Cervical cancer: role of imaging. *Abdom Imaging*. 1997;22:208–215.
5. Scheidler J, Hricak H, Yu KK. Radiological evaluation of lymph node metastases in patients with cervical cancer: a meta-analysis. *JAMA*. 1997;278:1096–1101.
6. Grigsby PW, Dehdashti F, Siegel BA. FDG-PET evaluation of carcinoma of the cervix. *Clin Positron Imaging*. 1999;2:105–109.
7. Grigsby PW, Siegel BA, Dehdashti F. Lymph node staging by positron emission tomography in patients with carcinoma of the cervix. *J Clin Oncol*. 2001;19:3745–3749.
8. Rose P, Adler L, Rodriguez M, et al. Positron emission tomography for evaluating para-aortic nodal metastasis in locally advanced cervical cancer before surgical staging: a surgicopathologic study. *J Clin Oncol*. 1999;17:41–45.
9. Sugawara Y, Eisbruch A, Kosuda S, et al. Evaluation of FDG-PET in patients with cervical cancer. *J Nucl Med*. 1999;40:1125–1131.
10. Reinhardt M, Ehrhrit-Braun C, Vogelgesang D, et al. Metastatic lymph nodes in patients with cervical cancer: detection with MR imaging and FDG PET. *Radiology*. 2001;218:776–782.
11. Narayan K, Hicks RJ, Jobling T, et al. A comparison of MRI and PET scanning in surgically staged loco-regionally advanced cervical cancer: potential impact on treatment. *Int J Gynecol Cancer*. 2001;11:263–271.
12. Miller TR, Grigsby PW. Measurement of tumor volume by PET to evaluate prognosis in patients with advanced cervical cancer treated by radiation therapy. *Int J Radiat Oncol Biol Phys*. 2002;53:353–359.
13. Morris M, Eifel PJ, Lu J, et al. Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aortic radiation for high-risk cervical cancer. *N Engl J Med*. 1999;340:1137–1143.
14. Xu M, Luk W, Cutler P, et al. Local threshold for segmented attenuation correction of PET imaging of the thorax. *IEEE Trans Nucl Sci*. 1994;41:1532–1537.
15. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging*. 1994;13:601–609.
16. Armitage P, Berry G. *Statistical Methods in Medical Research*. 2nd ed. Oxford, U.K.: Blackwell Scientific Publications; 1987:428–433.
17. Ingelfinger JA, Mosteller F, Thibodeau LA, et al. *Biostatistics in Clinical Medicine*. New York, NY: Macmillan Publishing Company; Inc.; 1983:28–30.
18. Glantz S. *Primer of Biostatistics*. New York, NY: McGraw-Hill; 1997:314–317.
19. Bentzen SM, Thames HD. Tumor volume and local control probability: clinical data and radiobiological interpretations. *Int J Radiat Oncol Biol Phys*. 1996;36:247–251.
20. DiChiro G, Brooks RA. PET quantitation: blessing and curse. *J Nucl Med*. 1988;29:1603–1604.
21. Green-Hernandez C, Singleton JK, Aronson DZ. *Primary Care Pediatrics*. Philadelphia, PA: Lippincott; 1996:97–98.