Risk Stratification of Gallbladder Polyps (1–2 cm) for Surgical Intervention with 18F-FDG PET/CT

Jaehoon Lee1, Mijin Yun1,2, Kyoung-Sik Kim3, Jong-Doo Lee1, and Chun K. Kim2

1Division of Nuclear Medicine, Department of Diagnostic Radiology, Yonsei University College of Medicine, Seoul, Korea; 2Division of Nuclear Medicine and Molecular Imaging, Department of Radiology, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts; and 3Department of General Surgery, Yonsei University College of Medicine, Seoul, Korea

We assessed the value of 18F-FDG uptake in the gallbladder polyp (GP) in risk stratification for surgical intervention and the optimal cutoff level of the parameters derived from GP 18F-FDG uptake for differentiating malignant from benign etiologies in a select, homogeneous group of patients with 1- to 2-cm GPs. Methods: Fifty patients with 1- to 2-cm GPs incidentally found on the CT portion of PET/CT were retrospectively analyzed. All patients had histologic diagnoses. GP 18F-FDG activity was visually scored positive (≥ liver) or negative (< liver). Maximal standardized uptake value of the GP (SUVgp) and ratio of SUVgp to mean SUV of the liver (GP/L ratio) were also measured. Univariate and multivariate logistic regression analyses were performed to determine the utility of patient and clinical variables—that is, sex, age, gallstone, polyp size, and three 18F-FDG-related parameters in risk stratification. Results: Twenty GPs were classified as malignant and 30 as benign. Multivariate analyses showed that the age and all parameters (visual criteria, SUVgp, and GP/L) related to 18F-FDG uptake were significant risk factors, with the GP/L being the most significant. The sex, size of GPs, and presence of concurrent gallstones were found to be insignificant. Conclusion: 18F-FDG uptake in a GP is a strong risk factor that can be used to determine the necessity of surgical intervention more effectively than other known risk factors. However, all criteria derived from 18F-FDG uptake presented in this series may be applicable to the assessment of 1- to 2-cm GPs.

Key Words: gallbladder polyp; gallbladder cancer; polypoid lesion of gallbladder; PET; fluorodeoxyglucose

DOI: 10.2967/jnumed.111.093948

Increased use and improved quality of ultrasonography has resulted in a significant increase in the prevalence of gallbladder polyps (GPs). The prevalence of GP was reported to be 5.6% in a large series with a sample size of 194,767 asymptomatic adults (1). Because the prognosis of gallbladder cancer is extremely poor (2), accurate differentiation between the malignant and benign GPs is essential. A GP size of 1 cm or greater is considered an important risk factor for malignancy requiring surgical intervention (3–5). Other factors used for risk stratification, either for differentiating malignant from benign lesions or for differentiating neoplastic from nonneoplastic lesions, include age, the number and shape of the GP, and the presence or absence of concurrent gallstone (1,5–9). Some features on high-resolution ultrasonography, endoscopic ultrasonography, and CT have been used to differentiate malignant from benign gallbladder lesions (10) or to differentiate neoplastic from nonneoplastic polyps (11–13). Although some investigators reported excellent results (10), significant uncertainty remains in the ability to differentiate benign from malignant GPs (14) or even in the ability to differentiate neoplastic from nonneoplastic polyps (15).

The utility of PET using 18F-FDG for differentiating malignant from benign gallbladder diseases has been reported (16–18). However, these investigations were limited because of the relatively small sample size and heterogeneous populations. For example, some patients had GPs whereas others had gallbladder wall thickening without GP. In addition, these investigations were performed using single-modality PET systems rather than PET/CT. Identification of relatively hypometabolic GP may be difficult on PET alone, potentially leading to errors in measurement of true 18F-FDG uptake value.

The purpose of this study was to assess the value of increased 18F-FDG uptake in the GP in risk stratification for surgical intervention and the optimal cutoff level of the parameters derived from the 18F-FDG uptake in a select, homogeneous group of patients with GPs. Specifically, patients with diffuse gallbladder disease or gallbladder wall thickening without GPs and those with either high or low pretest probability of malignancy, for example, GPs of 2 cm or greater or less than 1 cm, respectively, were excluded.

MATERIALS AND METHODS

Patients

After obtaining approval from the institutional review board, we retrospectively reviewed the medical records of patients who were incidentally found to have GPs on the diagnostic CT portion
of $^{18}$F-FDG PET/CT studies and subsequently underwent cholecystectomy in our institution over a 4-y period. All $^{18}$F-FDG studies were performed for indications unrelated to gallbladder disease. Patients with diffuse gallbladder disease or wall thickening without GPs were not included. Among patients with GPs, the following 3 groups of patients in whom the pretest probability of malignancy is considered either very high or very low were additionally excluded: patients with radiologic findings already highly suggestive of malignancy, for example, local invasion or metastasis in the adjacent organs or pathologic lymphadenopathy; patients with polyps larger than 2 cm in the greatest diameter; and patients with polyps smaller than 1 cm in the greatest diameter. After exclusion, there were 55 patients with 1- to 2-cm polyps. Of these, 5 patients elected to have clinical follow-up and 50 patients had surgery. These 50 polyps in 50 patients were available for evaluation. There were 26 women and 24 men, with a mean age of 59.5 y (age range, 34–79 y). The indication for PET/CT included staging or restaging lung cancer in 5 patients, colorectal cancer in 4 patients, breast cancers in 3 patients, lymphoma in 2 patients, other known or suspected malignancies in 8 patients, and cancer screening in 28 patients.

Final diagnosis was made by histology in all patients and classified into 2 categories—malignant or benign. Adenomas containing high-grade dysplasia or focal malignant transformation were categorized as malignant for the purpose of this investigation because surgery is indicated for such adenomas.

**Imaging Procedures**

All patients fasted for at least 6 h before the study and rested for at least 1 h before the PET/CT scan. Blood glucose concentration was measured and confirmed to be less than 140 mg/dL before scanning. Approximately 5.5 MBq of $^{18}$F-FDG per kilogram of body weight was administered intravenously, and the duration of the uptake phase was 60 min. Examinations were performed using a Biograph TruePoint 40 PET/CT scanner (Siemens Medical Systems, CTI), with an axial field of view of 21.6 cm and a spatial resolution of 4.2 mm in full width at half maximum at 1 cm from the center. A low-dose CT scan for attenuation correction was first obtained, immediately followed by emission imaging from the neck to mid thigh in 3-dimensional mode at 3 min per bed position. The patient then underwent a diagnostic CT scan with intravenous contrast. PET data were reconstructed iteratively using an ordered-subset expectation maximization algorithm, with the low-dose CT datasets used for attenuation correction.

**Image Interpretation and Analysis**

Visual grading of $^{18}$F-FDG activity in GPs was performed without knowledge of histology; $^{18}$F-FDG activity was scored positive if similar to or greater than liver parenchymal activity or visually clearly discernible and negative if lower or visually indiscernible [Fig. 1](#Fig1). For semiquantitative analysis, a region of interest (ROI) was placed over the identified GP. For GPs visualized on PET, the ROI was placed over the entire $^{18}$F-FDG–avid lesion on the axial PET images. However, if an accurate placement of the ROI on PET was deemed difficult because of little activity or no clearly discernible activity in the GP, the CT portion of the PET/CT study was used to place the ROI. The maximal standardized uptake value (SUV) in the GP (SUVgp) was recorded for each ROI. Mean SUV of the liver was obtained from an ROI placed over an area of homogeneous activity in the right lobe. Care was taken to avoid the central area of large vascular structures and any areas of increased $^{18}$F-FDG uptake that might represent tumor. The ratio of SUVgp to mean SUV of the liver (GP/L ratio) was also calculated.

**Statistical Analysis**

Statistical analyses were performed using MedCalc statistical software (version 9.6.4.0; MedCalc Software). An independent Student $t$ test or Mann–Whitney test was used to compare the 4 continuous independent variables—that is, the age of patients, GP size, SUVgp, and GP/L ratio—between the malignant and benign groups. Then the cutoff values of these variables providing the best separation between malignant and benign GPs were obtained using the receiver-operating-characteristic analyses and used to dichotomize the data. The diagnostic performance of these 4 cutoff values and of 3 additional categoric variables (sex, gallstone, and visual grading of $^{18}$F-FDG uptake in GPs) in differentiating malignant from benign GPs was assessed by univariate logistic regression analysis.

Independent variables with a $P$ value of less than 0.25 identified from the univariate analysis were included in the multivariate logistic regression analysis. Of the 3 variables derived from the $^{18}$F-FDG uptake in the GPs (visual grading, SUVgp, and GP/L ratio), only 1 variable that was associated with the lowest $P$ value and highest odds ratio was chosen to be included in the multivariate analysis. In addition, McNemar tests were performed to compare the performances of dichotomized visual, SUVgp, and GP/L ratio criteria. The results were considered statistically significant if the $P$ value from the multivariate analysis was less than 0.05 and the 95% confidence interval of the odds ratio did not include 1.

**RESULTS**

Twenty neoplastic polyps were categorized as malignant (16 adenocarcinomas, 2 adenomas with focal malignant transformation, and 2 adenomas with high-grade dysplasia). There were 30 benign polyps, including 26 nonneoplastic (12 inflammatory polyps, 9 cholesterol polyps, 3...
adenomyomatosis, and 2 papillary epithelial hyperplasias) and 4 neoplastic lesions (4 adenomas).

**Table 1**  Table 1 shows the comparison of the mean or median values of the continuous variables—that is, age, polyp size, SUVgp, and GP/L ratio—between malignant and benign lesions. There was no statistically significant difference in size between malignant and benign polyps, most likely because of the narrow range (1–2 cm) selected for this investigation. However, despite the narrow size range, there was a statistically significant difference in the age, SUVgp, and GP/L ratio between malignant and benign polyps. The GP/L ratio yielded the lowest $P$ value, 0.0001.

The receiver-operating-characteristic analysis yielded an age of 63 y, size of 1.2 cm, SUVgp of 2.1, and GP/L ratio of 1.14 to be the best cutoffs for each continuous variable, with an area under the curve of 0.750, 0.642, 0.772, and 0.822, respectively. Table 2 summarizes the results of univariate analyses of the 7 variables and those of multivariate analyses of 3 selected variables. Of the 7 variables tested, the sex and presence of concurrent gallstone did not show any trend. The cutoff GP size of 1.2 cm yielded a marginal $P$ value of 0.052, which approached statistical significance; this variable, therefore, was included in the multivariate logistic regression analysis. The cutoff age of 63 y, visual $^{18}$F-FDG uptake, SUVgp, and GP/L ratio were found to be significant by univariate analyses. Of the 3 variables related to $^{18}$F-FDG uptake, the GP/L ratio was associated with the lowest $P$ value, 0.00001, and the highest odds ratio, 36.8, and therefore chosen to be included in the multivariate analysis. On multivariate analysis including the age of 63 y, polyp size of 1.2 cm, and GP/L ratio, the GP/L ratio remained the most significant, followed by the age. The statistical significance of size deteriorated on the multivariate analysis.

**Table 2**  Table 3 shows the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy using the cutoff values of the 3 parameters derived from $^{18}$F-FDG uptake in the GPs. The GP/L ratio again yielded the highest values. However, the McNemar test yielded no significant difference in performance between any pair of these 3 variables.

The retrospectively determined cutoff GP/L ratio of 1.14 yielded false-negative results in 3 (1 adenocarcinoma and 2 adenomas with high-grade dysplasia) of the 20 patients with malignancy and false-positive results in 4 (3 of the 4 tubular adenomas and 1 of the 12 inflammatory polyps) of the 30 patients with benign GPs. None of the cases of adenomyomatosis ($n = 3$), cholesterol polyps ($n = 9$), or papillary epithelial hyperplasia ($n = 2$) were falsely positive.

**DISCUSSION**

GPs consist of true neoplasms, such as benign adenomas and adenocarcinomas, and nonneoplastic polyps, such as cholesterol polyps, inflammatory polyps, and adenomyomatous hyperplasia. Of the benign GPs, the cholesterol polyp is reported to be the most common type ($5,14,19–21$). However, the most common benign GP in our patients included in this study was the inflammatory polyp associated with cholecystitis (40%; 12/30). The relatively lower number of cholesterol polyps (30%; 9/30), compared with inflammatory polyps, in our report might have partly resulted from our selection criteria (i.e., most cholesterol polyps were smaller than 1 cm, and many of them were likely excluded).

Most malignant GPs were reported to be larger than 1 cm, with a sensitivity of 88%, yet the PPV of the 1-cm cutoff size for malignancy ranges widely from 31% to 78% (4,22). Of the 50 lesions of 1–2 cm included in our study, malignancy was found in 18 (16 gallbladder cancers and 2 adenomas with focal malignant transformation), translating to a PPV of 36% for the 1- to 2-cm size criterion for malignancy. The PPV of the 1-cm or greater size criterion should have been somewhat higher, if GPs greater than 2 cm were included. Regardless, this result certainly appears within the range of the reported PPVs, indirectly indicating that our patient population is probably not different from others in the literature.

Patients with malignant GPs were significantly older than those with benign GPs (65.5 ± 10.5 y vs. 55.5 ± 10.4 y, $P = 0.002$). The 2 most common cutoff ages used for risk stratification in the literature are 50 y (7,9) and 60 y (4,5,8). We found the optimal cutoff in our 1- to 2-cm GPs to be 63 y, which is closer to the cutoff age of 60 y than 50 y. However, although our results confirm that age is one of the

### TABLE 1

Comparison of Mean or Median Values of 4 Continuous Variables Between Benign and Malignant GPs

<table>
<thead>
<tr>
<th>Variable</th>
<th>Benign ($n = 30$)</th>
<th>Malignant ($n = 20$)</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>55.5 ± 10.4</td>
<td>65.5 ± 10.5</td>
<td>0.0018*</td>
</tr>
<tr>
<td>Polyp size (cm)</td>
<td>1.35 ± 0.33</td>
<td>1.51 ± 0.32</td>
<td>0.1008*</td>
</tr>
<tr>
<td>SUVgp (g/mL)</td>
<td>1.63 ± 0.001</td>
<td>3.28 ± 0.001</td>
<td>0.00011</td>
</tr>
<tr>
<td>GP/L ratio</td>
<td>0.82 ± 0.001</td>
<td>1.49 ± 0.001</td>
<td>0.00011</td>
</tr>
</tbody>
</table>

* $P$ value by independent Student $t$ test.
† $P$ value by Mann–Whitney test.
significant risk factors, it was not as significant as variables derived from 18F-FDG uptake.

Accuracies of 18F-FDG PET for differentiating malignant from benign gallbladder disease reported in 3 prior studies range from 72% to 84% (16–18). However, these results are not generalizable. The number of patients was relatively small—that is, 32 patients in 1 study (18) and 16 patients each in 2 studies (16,17). The small number of patients is compounded further by the heterogeneous populations, either mixing those with diffuse thickening and those with polypoid lesion (17) or including those with relatively large tumors (16). Also, the size (17,18) or the presence or absence of benign polypoid lesions (18) was not described. The median SUVgp in benign and malignant GPs in our series was 1.63 and 3.28, respectively. These are considerably lower than the SUVs of 5.4 (benign) and 7.35 (malignant) reported by Nishiyama et al. (18). The average SUV of tumor lesions reported by Rodriguez et al. (17) was 4.1, also higher than our values, despite the fact that their unit was a mean SUV whereas ours was a maximum SUV. Therefore, the real difference between our values and theirs is probably even greater. We strongly believe that the relatively lower SUVs in our study resulted from the well-known partial-volume effect (23–25) because all our GPs were in the 1- to 2-cm range. Last, because none of the 3 investigations used a fused PET/CT system, potential errors could have been introduced in SUV measurement, especially when GPs could not be visually separated from surrounding tissue. Direct coregistration of PET and CT images should yield more accurate results both visually and quantitatively than mental coregistration of separately obtained PET and CT images.

The results from our 50 patients, all with polypoid lesions of diagnostically challenging size in terms of risk stratification (i.e., 1–2 cm), indeed show that 18F-FDG uptake in GPs is a powerful tool for risk stratification.

### TABLE 2
Correlation of Dichotomized Variables with Histopathologic Diagnoses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Benign (n = 30)</th>
<th>Malignant (n = 20)</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Odds ratio</td>
<td>95% confidence interval</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>15 (63)</td>
<td>9 (37)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>15 (58)</td>
<td>11 (42)</td>
<td>1.2</td>
<td>0.4–3.8</td>
</tr>
<tr>
<td>Age (cutoff, 63 y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 63</td>
<td>24 (80)</td>
<td>6 (20)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>&gt; 63</td>
<td>6 (30)</td>
<td>14 (70)</td>
<td>9.3</td>
<td>2.5–34.6</td>
</tr>
<tr>
<td>Gallstone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>25 (60)</td>
<td>17 (40)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5 (63)</td>
<td>3 (37)</td>
<td>0.9</td>
<td>0.2–4.2</td>
</tr>
<tr>
<td>Polyp size (cm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 1.2</td>
<td>16 (76)</td>
<td>5 (24)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>&gt; 1.2</td>
<td>14 (48)</td>
<td>15 (52)</td>
<td>3.4</td>
<td>1.0–11.9</td>
</tr>
<tr>
<td>Visual 18F-FDG uptake</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; liver</td>
<td>24 (86)</td>
<td>4 (14)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>≥ liver</td>
<td>6 (27)</td>
<td>16 (73)</td>
<td>16.0</td>
<td>3.9–65.8</td>
</tr>
<tr>
<td>SUVgp (g/mL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 2.1</td>
<td>23 (85)</td>
<td>4 (15)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>&gt; 2.1</td>
<td>7 (30)</td>
<td>16 (70)</td>
<td>13.1</td>
<td>3.3–52.5</td>
</tr>
<tr>
<td>GP/L ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 1.14</td>
<td>26 (90)</td>
<td>3 (10)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>&gt; 1.14</td>
<td>4 (19)</td>
<td>17 (81)</td>
<td>36.8</td>
<td>7.3–185.6</td>
</tr>
</tbody>
</table>

Data in parentheses are percentages.

### TABLE 3
Overall Performance of Cutoff Values of 3 Parameters Derived from 18F-FDG Uptake by GPs

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Accuracy</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual 18F-FDG uptake</td>
<td>80%</td>
<td>80%</td>
<td>73%</td>
<td>86%</td>
<td>80%</td>
<td>0.800</td>
</tr>
<tr>
<td>SUVgp (cutoff value, 2.1 g/mL)</td>
<td>80%</td>
<td>77%</td>
<td>70%</td>
<td>85%</td>
<td>78%</td>
<td>0.772</td>
</tr>
<tr>
<td>GP/L ratio (cutoff value, 1.14)</td>
<td>85%</td>
<td>87%</td>
<td>81%</td>
<td>90%</td>
<td>86%</td>
<td>0.822</td>
</tr>
</tbody>
</table>

AUC = area under curves by receiver-operating-characteristic analyses.
visual grading, cutoff SUVgp of 2.14, and cutoff GPL/L ratio of 1.14 yielded accuracy of 80%, 78%, and 84%, with an area under the ROC curve of 0.800, 0.772, and 0.822, respectively. Although there was no statistically significant difference in diagnostic performance among these 3 parameters, the SUVgp yielded the lowest values. Moreover, it is well known that variations in SUV may result from factors unrelated to the patient, such as body composition and habitus (29,30), length of uptake period (31), and blood glucose level (32), are problems common to all laboratories. Therefore, we believe that the GPL/L ratio or the visual criteria would be more generalizable and should be used by other laboratories.

Laparoscopic cholecystectomy is a relatively simple procedure. If this procedure is to be used in every patient with a GP, the value of PET would be limited despite its high NPV and overall efficacy. However, surgeons may feel that the results from PET, when combined with other available clinical information, are useful in decision making not only in terms of laparoscopic procedure versus open laparotomy but also wide excision versus limited resection if the latter is to be performed. The utility of PET in this context is currently under investigation.

This was a retrospective investigational study. Therefore, inherent bias in patient selection related to the retrospective nature might have been introduced. Nonetheless, some of the more obvious limitations existing in the previous reports were partially overcome by studying a larger, more homogeneous group of patients with GPs, by excluding those with gallbladder wall thickening only, and by the use of additional coregistered CT images. Therefore, we believe that the results from this study provide new insight as to the true value of 18F-FDG PET/CT for differentiating malignant from benign GPs.

CONCLUSION

Our data show that 18F-FDG uptake in GPs is a useful predictor in risk stratification for surgical intervention, perhaps more effective than many other known risk factors. The cutoff polyp-to-liver 18F-FDG uptake ratio of 1.14 and visual grading, compared with liver activity, were both associated with a reasonably high diagnostic efficacy, even when used in isolation. All criteria related to 18F-FDG uptake presented in this series were based on the GPs in the 1- to 2-cm range. Therefore, these criteria may be applicable in assessing similar-sized polyps only.

DISCLOSURE STATEMENT

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

ACKNOWLEDGMENT

This work was partially supported by a Korea Science and Engineering Foundation (KOSEF) grant from the Ministry of Education, Science, and Technology (MEST) (M20702010003-08N0201-00314) and a faculty grant of Yonsei University College of Medicine for 2011(6-2011-0165). No other potential conflict of interest relevant to this article was reported.

REFERENCES


Risk Stratification of Gallbladder Polyps (1–2 cm) for Surgical Intervention with $^{18}$F-FDG PET/CT

Jaehoon Lee, Mijin Yun, Kyong-Sik Kim, Jong-Doo Lee and Chun K. Kim

*J Nucl Med.*
Published online: February 7, 2012.
Doi: 10.2967/jnumed.111.093948

This article and updated information are available at:
http://jnm.snmjournals.org/content/early/2012/02/07/jnumed.111.093948

Information about reproducing figures, tables, or other portions of this article can be found online at:
http://jnm.snmjournals.org/site/misc/permission.xhtml

Information about subscriptions to JNM can be found at:
http://jnm.snmjournals.org/site/subscriptions/online.xhtml

*JNM* ahead of print articles have been peer reviewed and accepted for publication in *JNM*. They have not been copyedited, nor have they appeared in a print or online issue of the journal. Once the accepted manuscripts appear in the *JNM* ahead of print area, they will be prepared for print and online publication, which includes copyediting, typesetting, proofreading, and author review. This process may lead to differences between the accepted version of the manuscript and the final, published version.