Inter- and intraobserver consistency in assessing eligibility for bevacizumab (BVZ) in non-small-cell lung cancer (NSCLC) patients with centrally located tumors


1Department of Thoracic Oncology, Université de la Méditerranée—Assistance Publique Hôpitaux de Marseille, Marseille; Departments of 2Radiology and 3Medical Oncology, Institut Gustave Roussy, Villejuif; 4Department of Radiology, Centre Oscar Lambret, Lille; 5Department of Radiology, Centre François Baclesse, Caen; 6Department of Radiology, Centre Hospitalier Universitaire de Bordeaux, Bordeaux; 7Department of Medical Oncology, Centre Oscar Lambret, Lille; 8Department of Radiology, Centre Hospitalier Universitaire de Caen; 9Department of Radiology, Assistance Publique Hôpitaux de Paris, Paris; 10Department of Biostatistics, Assistance Publique Hôpitaux de Marseille, Marseille; 11Department of Respiratory Diseases, Centre Hospitalier Universitaire de Grenoble, Grenoble and 12Department of Respiratory Diseases, Centre Hospitalier Universitaire de Lille, Lille, France

Received 31 August 2009; revised 25 November 2009; accepted 26 November 2009

Background: Bevacizumab (BVZ) combined with platinum-based therapy is registered for first-line treatment of non-squamous non-small-cell lung cancer (NSCLC). Patients with centrally located tumors are stated ineligible for BVZ treatment. The goal of this study was to assess the consistency in evaluating eligibility of patients with central tumors for BVZ treatment.

Materials and methods: The study group was composed of 150 NSCLC patients with centrally located tumors. Eligibility for BVZ was assessed by chest computed tomography (CT) scan. Eligibility was assessed independently using CT images reviewed on workstations. Inter- and intraobserver variations on 50 randomly extracted patients were estimated through a statistical modeling (multiple correspondence analysis).

Results: Discordance in eligibility was found for 82 patients (55%). The interobserver strength of agreement was fair to moderate (average kappa = 0.40). Contrarily, the intraobserver strength of agreement was good to very good (average kappa = 0.74). At multivariate analysis, the risk of discrepancy was essentially related to the assessment of the contact between the tumor and the vessels (odds ratio = 13.3, 95% confidence interval 2.8–62.6, P = 0.001).

Conclusions: The consistency in evaluating eligibility of patients with central tumors for BVZ treatment is weak. The study group indicated more stringent criteria to help physicians in taking the best treatment choice that need however to be prospectively validated.

Key words: antiangiogenic drugs, bevacizumab, consistency, CT scan, non-small-cell lung cancer

introduction

Sustained angiogenesis is one of the six major hallmarks of cancer cells [1]. Consequently, during the last decade, several anticancer drugs targeting angiogenesis have been developed. Bevacizumab (BVZ) is an mAb that targets the vascular endothelial growth factor (VEGF) and showed activity in colon, breast, and non-small-cell lung cancer (NSCLC) [2–4]. However, BVZ-based therapy has been associated with bleeding adverse events in clinical trials across different indications. The most severe events occurred in a phase II trial of BVZ in patients with NSCLC with 9% of severe pulmonary hemorrhages (PH) leading to four deaths [5]. Beyond BVZ, severe PH associated with other antiangiogenic therapy based on VEGF or multitarget receptor tyrosine kinase inhibitors such as sunitinib, sorafenib, and vandetanib have been also reported indicating a class effect [6]. Regarding BVZ, these severe PH were subsequently associated with squamous cell histology [squamous cell carcinoma (SCC)] and disease location close to major blood vessels. Accordingly, patients with SCC and/or a centrally located tumor were excluded from BVZ-based therapies in subsequent clinical trials and now from the first-line treatment of NSCLC patients in the countries where the drug is registered in this indication [7]. However, several factors including baseline hemoptysis, tumor cavitation, tumor location, tumor size, etc. have been alternatively indicated to predict occurrence of PH with still a limited number of PH events (1.8%) precluding a formal multivariate analysis [8–12]. To date, the baseline...
factors to be considered to state (in)eligibility to BVZ appear therefore doubtful.

The methodology to define a mediastinal involvement by the tumor is not clear and suffers from the lack of specific recommendations, possibly translating into difficulties in determining patients’ eligibility for BVZ and, potentially, other antiangiogenic treatments. Inasmuch as noninvasive procedures must be preferred, the thoracic computed tomography (CT) scan now represents the cornerstone of this selection. On the basis of surgical series, some criteria have been proposed to predict the contact between the tumor and the adjacent mediastinal structures. Herman et al. [13] indicated that a contact >180° provides with a positive predictive value (PPV) for vascular invasion of 100%, while a CT evidence of an intraluminal tumor provides with a PPV for vascular involvement of only 65%. Glazer et al. [14] indicated that a tumor is unlikely to involve adjacent mediastinal structures if at least one of the following features is present (<3 cm of contact between the tumor and the mediastinum; <90° of circumferential contact with the aorta; and a visible mediastinal fat plane between the tumor and the mediastinal structure). However, a discrepancy between CT scan and postoperative findings was present in 21 of the 48 tumors in this series. In summary, the accuracy of the thoracic CT scan in defining the contact between the tumor and the adjacent mediastinal structures is low ranging from 56% to 89% in the most recent reviews [15, 16].

Therefore, determining the involvement of adjacent mediastinal structures by the tumor in NSCLC patients must be standardized, especially to preserve comparability across clinical trials and select NSCLC patients to receive BVZ or more widely antiangiogenic drugs that could lead to life-threatening PH. Then, the purpose of this study was to evaluate the factors that are now considered by oncologists and radiologists and the consistency among observers in assessing the involvement of mediastinal structures by NSCLC tumors to state BVZ (in)eligibility.

materials and methods

group study

Overall, 150 baseline chest CT scans from unselected stage IIIB/IV NSCLC chemonaive patients with centrally located tumors retrieved from eight academic hospitals’ databases were reviewed. A central lesion was defined as a lesion that had a distance of ≤2 cm between the edge of the tumor and the trachea, main, and lobar bronchi. No history of bleeding has come to the observers’ knowledge at the time of data collection.

CT scans

All the CT scans were assessed for quality of contrast enhancement and image quality. All the CT scans included in the study should have been carried out after i.v. administration of 2 ml/kg of iodine on multislice CT with at least four detector rows. Acquisition was carried out from 50 to 90 s after the onset of injection. Injection rate was 1–3 ml/s. Acquisition collimation depended on the machine, but the image reconstruction was obtained with a maximal slice thickness of 5 mm and contiguous images. Additional reformatted images in the coronal and sagittal planes were optional. CT images were reviewed on image interpretation workstations.

consensus regarding radiological criteria for BVZ eligibility

As a first step (Figure 1), the 150 CT studies were made available to a panel, including seven radiologists and seven oncologists or chest physicians, all

![Figure 1. Study design. CT, computed tomography.](image-url)

involved in lung cancer management. In addition, the panel reviewed all the exclusion criteria related to either the definition of a centrally located tumor or a tumor involving the mediastinum or the major blood vessels that have been considered along the ECOG4599, AVAiL and SAIL studies (i.e. especially a tumor that is fully contiguous with, surrounding, or extending into the lumen of a major blood vessel as pulmonary artery or superior vena cava). Accordingly, after training and discussion with an exploratory analysis of discordance and agreement rates in (in)eligibility between observers, the panel proposed the following methodology for thoracic CT scan analysis. The following criteria were considered for BVZ ineligibility: (i) central involvement by tumor but not by lymph node (except lymph node with extracapsular extension—Figure 2A); (ii) a contact >180° with a proximal artery (Figure 2B); and (iii) a partial involvement of a main bronchus or the trachea (defined by tumor partly or entirely encasing the tracheobronchial tree with or without tracheobronchial narrowing, tracheobronchial tree ended in a mass with or without consolidation, or localized intraluminal soft tissue density—Figure 2C). The following criteria were considered as needing additional imaging (especially multiplanar reconstruction) before stating BVZ ineligibility: (i) tumors with a contact with smaller vessels, atelectasis (Figure 2D), or (ii) cavitation (defined by the presence of air in any intrathoracic lesion).

The panel did not consider other (i.e. clinical) criteria in addition to radiological criteria to assess BVZ eligibility (patients were otherwise deemed eligible for the drug).

inter- and intraobserver consistency

As a second step (Figure 1), the inter- and intraobserver variations were determined 1 month later on 50 CT scans randomly extracted from the databases. A small panel consisting of three radiologists and three oncologists independently reviewed the 50 CT scans on two different sessions a half day apart. The decision on BVZ eligibility (yes/no) was independently determined on the basis of the predefined criteria.

Kappa statistics were carried out, as previously described [17], to determine the interobserver and intraobserver reliabilities of the BVZ eligibility. Kappa values of each possible pair of the six observers were calculated for interobserver reliability of the decision on BVZ eligibility in the first and second evaluation. Intraobserver reliability was assessed by comparing the decision on BVZ eligibility by each observer on the two assessments. The kappa coefficient of reliability provides a pairwise proportion of agreement between or among observers, corrected by chance. The kappa value was
results

assessment of BVZ eligibility by the panel

After assessment of the overall 150 CT scans by the whole panel, discordance in eligibility was found for 82 patients (55%), with only one discrepancy in the decision for 31 patients (20%). Conversely, agreement was reached in 68 cases (45%), with 59 patients deemed as ineligible and 9 patients deemed eligible for the drug. Analysis of variations showed significant differences among physicians independently of their specialty ($P < 0.05$).

assessment of interobserver consistency

Table 1 summarizes the results of the interobserver consistency. Overall, the strength of agreement was fair to moderate and was better between radiologists, than between oncologists and radiologists or between oncologists.

The kappa values of the two agreements were not statistically different when compared using the paired $t$-test ($P = 0.26$) meaning that agreements/disagreements between two observers remain relatively stable between the two assessments.

assessment of intraobserver consistency

Table 2 summarizes the results of the intraobserver consistency. Contrarily to the interobserver consistency, the strength of agreement was good to very good.

multiple correspondence analysis

Figure 3 shows the results of the MCA regarding first and second assessment of the eligibility by each observer. Each point represents an observer decision (eligibility versus ineligibility) at the first and the second assessment. The figure indicates a very close distance at each assessment, whatever the decision (eligibility or ineligibility), between the points representing the radiologists, while the distance between the points representing the oncologists are more scattered. In addition, distances are closest when deciding ineligibility (right part of the figure) than eligibility (left part of the figure).

multivariate analysis

The logistic regression model has been applied on the data of the first and the second assessment. On the first dataset, the risk of discrepancy in the decision of eligibility was significantly and independently higher when the observer should decide if a tumor involves the vessels or not (OR = 12.0, 95% CI 2.4–59.0, $P = 0.002$) or the trachea or main bronchus (OR = 10.8, 95% CI 1.5–77.4, $P = 0.018$). On the second dataset, the risk of discrepancy was essentially due to the assessment of the contact between the tumor and the vessels (OR = 13.3, 95% CI 2.8–62.6, $P = 0.001$).

discussion

BVZ has shown activity in colon cancer, breast cancer, and NSCLC and is now registered as first-line treatment in this late indication in combination with a platinum-based chemotherapy. However, BVZ has been associated with severe PH in patients with SCC, which has been accordingly excluded from the Food and Drug Administration and European Medicines Agency.
approvals of the drug. In addition, some other risk factors for severe pulmonary bleeding including centrally located tumors were considered as exclusion criteria in two phase III trials in NSCLC patients [2, 22]. Therefore, in clinical practice, identifying those patients in whom the risks outweigh the benefits associated with BVZ is challenging. The first reason relates to the benefit–risk ratio of the drug with excesses in ineligibility depriving the patients of the increase in response rate and survival benefit related to BVZ while excesses in eligibility exposing the patients to an unacceptable risk of death. The second reason relates to the lack of predictive factors of PH. The third challenging reason relates to the difficulties in determining the contact between the tumor and the mediastinal structures as attested by the results of the present study. Indeed, despite a settlement to analyze thoracic CT images with commonly accepted criteria, the strength of the agreement between trained observers was fair and disappointing. It is likely that these difficulties will be reproduced in daily practice and improvements in techniques and methodology are therefore needed.

Table 1. Interobserver consistency in assessing bevacizumab eligibility of non-small-cell lung cancer patients with centrally located tumors on CT scan during two different assessments

<table>
<thead>
<tr>
<th>Observers</th>
<th>First assessment</th>
<th>Second assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed agreement (%)</td>
<td>95% CI</td>
</tr>
<tr>
<td>O1 and O2</td>
<td>58</td>
<td>43.3–71.5</td>
</tr>
<tr>
<td>O1 and R1</td>
<td>60</td>
<td>45.2–73.2</td>
</tr>
<tr>
<td>O1 and R2</td>
<td>64</td>
<td>49.1–76.7</td>
</tr>
<tr>
<td>O1 and R3</td>
<td>60</td>
<td>45.2–73.3</td>
</tr>
<tr>
<td>O1 and O3</td>
<td>72</td>
<td>57.3–83.3</td>
</tr>
<tr>
<td>O2 and R1</td>
<td>74</td>
<td>59.4–84.9</td>
</tr>
<tr>
<td>O2 and R2</td>
<td>70</td>
<td>55.2–81.7</td>
</tr>
<tr>
<td>O2 and R3</td>
<td>78</td>
<td>63.7–88.0</td>
</tr>
<tr>
<td>O2 and O3</td>
<td>78</td>
<td>63.7–88.0</td>
</tr>
<tr>
<td>R1 and R2</td>
<td>84</td>
<td>70.3–92.4</td>
</tr>
<tr>
<td>R1 and R3</td>
<td>88</td>
<td>75.0–95.0</td>
</tr>
<tr>
<td>R1 and O3</td>
<td>72</td>
<td>57.3–83.3</td>
</tr>
<tr>
<td>R2 and R3</td>
<td>80</td>
<td>65.9–89.5</td>
</tr>
<tr>
<td>R2 and O3</td>
<td>76</td>
<td>61.5–86.5</td>
</tr>
<tr>
<td>Average kappa</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

CT, computed tomography; CI, confidence interval; O, oncologist; R, radiologist.

Table 2. Intraobserver consistency in assessing bevacizumab eligibility of non-small-cell lung cancer patients with centrally located tumors on CT scan during two different assessments

<table>
<thead>
<tr>
<th>Observers</th>
<th>First assessment observed agreement (%)</th>
<th>95% CI</th>
<th>Kappa value 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>80</td>
<td>65.9–89.5</td>
<td>0.60</td>
</tr>
<tr>
<td>O2</td>
<td>86</td>
<td>72.6–93.7</td>
<td>0.55</td>
</tr>
<tr>
<td>O3</td>
<td>92</td>
<td>79.9–97.4</td>
<td>0.81</td>
</tr>
<tr>
<td>R1</td>
<td>90</td>
<td>77.4–96.3</td>
<td>0.74</td>
</tr>
<tr>
<td>R2</td>
<td>90</td>
<td>77.4–96.3</td>
<td>0.78</td>
</tr>
<tr>
<td>R3</td>
<td>98</td>
<td>88.0–99.9</td>
<td>0.96</td>
</tr>
<tr>
<td>Average kappa</td>
<td>-</td>
<td>-</td>
<td>0.74</td>
</tr>
</tbody>
</table>

CT, computed tomography; CI, confidence interval; O, oncologist; R, radiologist.

Figure 3. Results of the multiple correspondence analysis (MCA) regarding first and second assessment of the bevacizumab eligibility by each observer (circles represent radiologists and squares represent oncologists). Each point represents an observer decision, with eligibility colored in green and ineligibility colored in red, at the first (full filling) and the second (hachured filling) assessment.

The inconsistency of the radiological measurements of lung tumors is of concern. Indeed, Gierada et al. [23] reported a moderate agreement (0.58–0.64) among radiologists on the interpretation of pulmonary findings at low-dose CT screening examinations for lung cancer. Erasmus et al. [24] reported a significant difference among readers when evaluating lung tumor size that can lead to an incorrect interpretation of tumor response. A great concern is paid to the assessment of the connection between tumor and mediastinal structures especially when a surgical lung resection is planned. However, the accuracy of the CT scan to predict mediastinal structures and especially vascular invasion is poor ranging from 56% to
Surgical invasive procedures are unconceivable for the daily management of advanced NSCLC. For that reason, recent advances have taken advantage of the multiplanar reconstruction capability of multidetector CT, while there were in the literature few attempts to evaluate the clinical impact of such approach of anatomical details, namely, whether thinner slices and multiplanar reformatted images should allow a better delineation of the tumor contours and mediastinal or vascular involvement is unknown. Higashino et al. [25] using 1.25-mm-thick slices and reformatted images in 60 patients with NSCLC found that the overall reading time decreased as compared with cases evaluated in the axial plane only and demonstrated that reformatted images were superior to identify interlobar extent and chest wall invasion. There was unfortunately no comment about the mediastinal and/or vascular encasement in this study. In the future, these new techniques combined with computer-aided assessments might improve interobserver consistency of radiological assessments as indicated by a recent work [26].

The low rate of severe PH makes prospective studies on radiological factors associated with the risk of severe PH difficult, and as stated before, solely retrospective data are available. The final results from the two observational studies, the ARIES study in North America and the SAIL study in Europe, are awaited, but to date, few PH have been reported without identifying new clinical or radiological factor [10, 11]. This lack of certitude may explain in part the divergence between the observers, leading to the lack of consistency especially between oncologists as demonstrated by the MCA. A less homogeneous training and a lower experience in medical imaging might explain the greater heterogeneity for oncologists and chest physicians as assessed by the MCA when compared with radiologists.

In conclusion, besides the scientific uncertainties, the present study indicates that NSCLC patients’ eligibility for BVZ suffers from a great interobserver inconsistency, greater between oncologists than between radiologists. A strong collaboration between them is therefore mandatory as well as international standard guidelines to better drive the decision in daily practice to combine or not BVZ in otherwise eligible patients.

funding

Roche, France.

acknowledgements

The study has been presented at the European Society of Medical Oncology congress in September 2008.

references

12. Fischbach NA, Spigel D, Brahmer J et al. Preliminary safety and effectiveness of bevacizumab (Bv) based treatment in subpopulations of patients (pts) with non–small cell lung cancer (NSCLC) from the ARIES study; a bevacizumab (Bv) treatment observational cohort study (OCS). J Clin Oncol 2009; 27(15s): (Abstr 8040).