Multi-modal magnetic resonance imaging-based grading analysis for gliomas by integrating radiomics and deep features

Zhenyuan Ning[1,2#], Jiaxiu Luo[1,2#], Qing Xiao[1,2], Longmei Cai[3], Yuting Chen[3], Xiaohui Yu[1,2], Jian Wang[1,2,3], Yu Zhang[1,2]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou, China; [2]Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, China; [3]Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, China

Contributions: (I) Conception and design: Y Zhang, J Wang, Z Ning, J Luo; (II) Administrative support: Y Zhang; (III) Provision of study materials or patients: Z Ning, J Luo; (IV) Collection and assembly of data: J Wang; (V) Data analysis and interpretation: Z Ning, J Luo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Yu Zhang, PhD. School of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong, 510515, China. Email: yuzhang@smu.edu.cn; Jian Wang, PhD. School of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong, 510515, China. Email: wangj_gz1981@163.com.

Background: To investigate the feasibility of integrating global radiomics and local deep features based on multi-modal magnetic resonance imaging (MRI) for developing a noninvasive glioma grading model.

Methods: In this study, 567 patients [211 patients with glioblastomas (GBMs) and 356 patients with low-grade gliomas (LGGs)] between May 2006 and September 2018, were enrolled and divided into training (n=186), validation (n=47), and testing cohorts (n=334), respectively. All patients underwent postcontrast enhanced T1-weighted and T2 fluid-attenuated inversion recovery MRI scanning. Radiomics and deep features (trained by 8,510 3D patches) were extracted to quantify the global and local information of gliomas, respectively. A kernel fusion-based support vector machine (SVM) classifier was used to integrate these multi-modal features for grading gliomas. The performance of the grading model was assessed using the area under receiver operating curve (AUC), sensitivity, specificity, Delong test, and *t*-test.

Results: The AUC, sensitivity, and specificity of the model based on combination of radiomics and deep features were 0.94 [95% confidence interval (CI): 0.85, 0.99], 86% (95% CI: 64%, 97%), and 92% (95% CI: 75%, 99%), respectively, for the validation cohort; and 0.88 (95% CI: 0.84, 0.91), 88% (95% CI: 80%, 93%), and 81% (95% CI: 76%, 86%), respectively, for the independent testing cohort from a local hospital. The developed model outperformed the models based only on either radiomics or deep features (Delong test, both of P<0.001), and was also comparable to the clinical radiologists.

Conclusions: This study demonstrated the feasibility of integrating multi-modal MRI radiomics and deep features to develop a promising noninvasive grading model for gliomas.

Keywords: Glioma grading; integrative analysis; radiomics; deep learning; artificial intelligence (AI)

Introduction

Gliomas are the most common tumors of the central nervous system, accounting for 80% of all malignant tumors in the brain (1). In accordance with the World Health Organization criteria, gliomas are categorized into low-grade gliomas (LGGs) and glioblastomas (GBMs) in terms of histopathological findings (2,3). Preoperative glioma grading is important and meaningful for treatment decision

Page 2 of 12

Ning et al. Gliomas grading using radiomics and deep features

and prognosis analysis (3-5). Histopathological diagnosis after the biopsy is the golden standard for glioma grading. However, its invasiveness may introduce discomfort to the patients (6-8). Accordingly, an accurate and noninvasive model is helpful for the preoperative grading of gliomas.

Radiomics provides an efficient and feasible analysis for constructing a noninvasive model based on high-throughput feature extraction. It has been used in various clinical tasks, such as disease detection, diagnosis, and prognosis analysis (9-11). Several studies have developed radiomics models for grading gliomas by extracting global radiomics features from entire regions of interest (ROIs) or volumes of interest (VOIs) on magnetic resonance imaging (MRI) sequences, such as contrast enhanced T1-weighted (T1ce) and T2 fluid-attenuated inversion recovery (T2 FLAIR) (12-14). However, these global radiomics features may lose local and glioma-specific information. Currently, deep learning-based methods have shown promising performance in medical image analysis (15-17) and have also been used for glioma grading (18,19). The advantage of deep learning-based approaches is that they can learn deep features automatically, instead of extracting hand-crafted radiomics features (20-22). To generate sufficient data for model training, many deep learning-based methods have utilized a patch-based strategy for glioma grading (23-25). Compared with the "global" radiomics features extracted from whole VOIs, the deep features extracted from patches can be regarded as the "local" features of the VOIs. Naturally, it raises the idea that whether the combination of "local" and "global" features in the glioma grading model will outperform the models based on individual "local" or "global" features.

In this work, we aimed to investigate the feasibility of integrating global radiomics and local deep features based on multi-modal MR images for developing a noninvasive glioma grading model. First, radiomics and deep features were extracted to quantify the global and local information of gliomas, respectively. Then, a kernel fusion-based support vector machine (SVM) classifier was used to integrate these multi-modal features for grading gliomas. The performance of the grading model was assessed using the area under receiver operating curve (AUC), sensitivity, specificity, Delong test, and *t*-test. The results showed that our proposed model outperformed the models based only on either radiomics or deep features, and was also comparable to the clinical radiologists. We present the following article in accordance with the TRIPOD reporting checklist (available at http://dx.doi.org/10.21037/atm-20-4076).

Methods

Patient cohorts

This retrospective study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Institutional Review Board of Nanfang Hospital (Guangzhou, Guangdong, China; ID: NFEC-2020-251) and individual consent for this retrospective study was waived. Two datasets were collected in this study: an open dataset from The Cancer Imaging Archive, TCIA (http://www.cancerimagingarchive.net/) and the other from a local hospital. Totally, 567 patients with gliomas recorded from May 2006 to September 2018 were enrolled.

The open dataset consisted of 233 patients, including 106 patients with GBMs and 127 patients with LGGs. We randomly and equiprobably divided the dataset into the training cohort (n=186) and the validation cohort (n=47) with the ratio of 4:1, to train and select the parameters of model, respectively. For evaluating the developed model, an independent cohort from Nanfang Hospital was recruited as the external testing cohort, which comprised 334 patients, including 105 patients with GBMs and 229 patients with LGGs. All patients were pathologically confirmed as gliomas. The inclusion criteria were as follows: preoperative MR image data; available T2 FLAIR and T1ce images; high image quality without significant head motion or artifacts; and available histological grading information.

MRI acquisition and VOI segmentation

All patients underwent multi-institutional routine clinically preoperative MRI scanning, including T1ce and T2 FLAIR. For the TCIA cohort, images were acquired by using the magnetic field of 1.5T and 3T MRI systems from multiple institutions. Imaging parameters were as follows: repetition time and echo time, 5–11,000 and 0–155 msec, respectively; slice thickness, 2.5–6 mm; percentage phase field of view, 70–100%; flip angle, 90° or 150°; matrix size, 256×256 or 512×512.

For the independent testing cohort, MR images were acquired from one of three MR scanners: a 1.5T MR scanner (Achieva, Philips Healthcare, Best, The Netherlands), with a repetition time of 214 msec, echo time of 4.6 msec, slice thickness of 6 mm, flip angle of 80°, percentage phase field of view of 82%, and matrix of 512×512; a 3T MR scanner (Signa, GE Healthcare, Milwaukee, Wis, USA), with a repetition time of 600 msec, echo time of 17 msec, slice thickness of 5 mm, flip angle of

90°, percentage phase field of view of 75%, and matrix of 256×256; a 1.5T MR scanner (Avanto, Siemens Healthcare, Erlangen, Germany) with a repetition time of 663 msec, echo time of 17 msec, slice thickness of 5 mm, flip angle of 90°, percentage phase field of view of 100%, and matrix of 512×512.

The VOIs (including the whole glioma, peritumoral edema, and necrotic regions) were manually delineated on T2 FLAIR images by using ITK-SNAP 3.6 (ITK-SNAP 3.x Team, www.itksnap.org) by a radiologist with 10 years of experience. The contours of VOIs were copied to T1ce images that were aligned with T2 FLAIR images via rigid registration (*Figure 1A*).

Extraction of global radiomics features

As shown in *Figure 1B*, radiomics features were extracted based on entire VOIs, including non-texture and texture features (26). Non-texture features included the size, solidity, volume, and eccentricity of VOIs. Since different scan parameters and irrelevant information in the images would influence the texture feature extraction, several image preprocessing operators were performed to normalize all MR images, including wavelet band-pass filtering [weighted ratio: (1/2, 2/3, 1, 3/2, 2)], isotropic resampling [resampling size: (in-pR, 1, 2, 3, 4, and 5)], and gray level quantization {algorithm: (Equal, Lloyd); number of gray level: [8, 16, 32, 64]} (26) (see details in Appendix 1). Then the first-order (global), second-order (gray-level co-occurrence matrix), and high-order (gray-level run-length matrix, gray-level zone size matrix, and neighborhood gray-tone difference matrix) features were extracted.

Extraction of local deep features

We developed a convolutional neural network (CNN) model to extract deep features and used the pathologically confirmed grade as a reference standard (*Figure 1C*). The input of the CNN was non-overlapping 3D patches with a size of 24×24×24. Since gliomas are of various sizes, the choice of 24×24×24 is to ensure that the relatively small gliomas can also generate sufficient patches to train the network. Data augmentation was performed by image random rotation, translation, and zooming. Finally, a total of 9,940 3D patches were extracted from the TCIA cohort, in which 8,510 and 1,430 patches were extracted in training and validation cohorts, respectively. The designed CNN structure contained three convolutional blocks and each
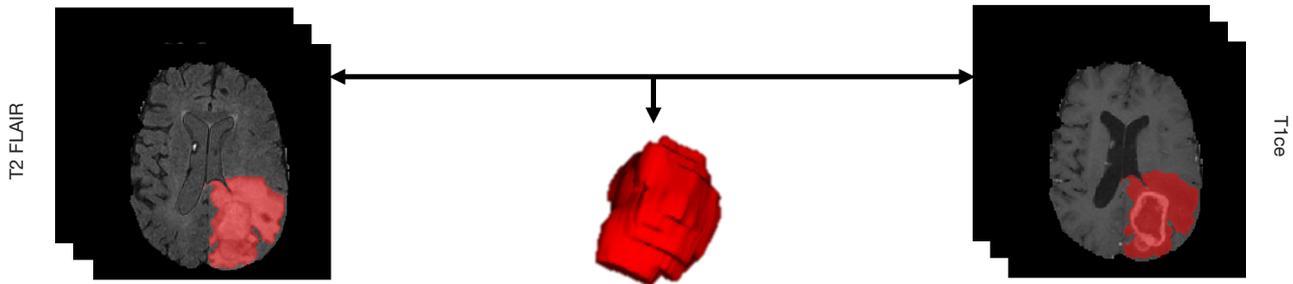
of the first two blocks was followed by an average pooling layer (see details in Figure S1). For each convolutional layer, kernel size was chosen as 2×2×2 with a stride of 1 and a padding size of 1, which could capture highly relevant edge information and involve detailed local textures. The kernel number of convolutional layers in three convolutional blocks were set to 32, 64, and 64, respectively. To prevent overfitting, a dropout operator with a rate of 0.2 or 0.3 was plugged into the 2×2×2 average pooling layer and the last convolutional layer. The crucial parameters of the CNN structure were experimentally tuned by internal validation cohort and will be discussed later. At the end of the network, two fully connected layers with a sigmoid activation function were used to grade gliomas. In the training step, the network was optimized using RMSprop optimizer (27) and the weights and bias were updated by the minimal batch which contained 16 patches. A predefined number of epochs was 250, and the training would been stopped when the network showed no significant performance improvement on the internal validation cohort. The learning rate was experimentally set to 0.00001.

For deep feature extraction, the feature maps output by the last convolutional layer was reshaped to a vector as a deep feature vector. Consequently, the proposed CNN could extract a feature vector for each patch. In clinical application, the objects of study were patients rather than patches. Meanwhile, the patches from the same patient might be classified by the network into different categories and confuse decision-making. To overcome this problem, we employed an average pooling strategy (28) to integrate the deep features of patches sampled from the same patient. This average pooling performed an average operation on each corresponding element in the feature vectors of all patches from the same patient, and obtained a final deep feature vector for each patient.
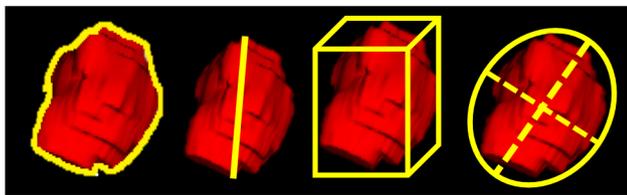
Feature reduction

To reduce redundancy among features and suppress overfitting, relief algorithm (29) was used to select the features with the best distinguishing power. According to relief algorithm, *k* features ranked by the importance were employed for classification (see Appendix 2). The relief algorithm was simultaneously applied to the four feature sets (i.e., T1ce radiomics, T1ce deep, T2 FLAIR radiomics, and T2 FLAIR deep features), and parameter *k* was determined in terms of the average value of the four AUCs for grading glioma. In addition, we also compared two classical feature
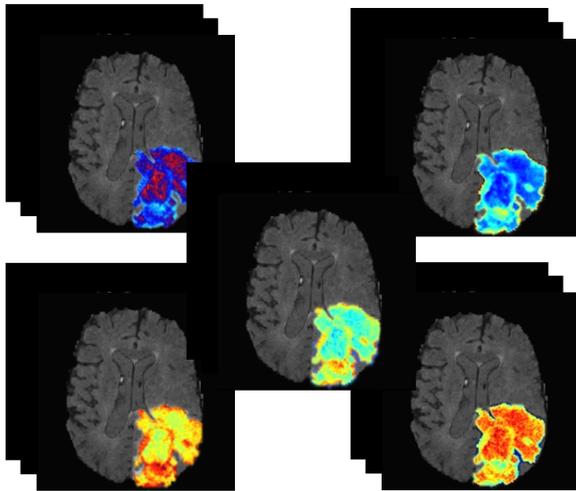
Page 4 of 12

Ning et al. Gliomas grading using radiomics and deep features
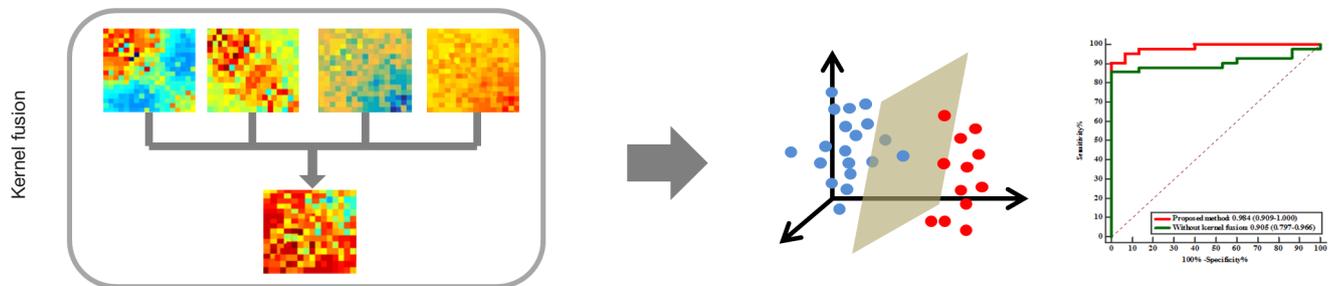
Figure 1 The flowchart of the proposed integrative framework. It included four steps, namely, (A) tumor imaging and segmentation, (B) radiomics feature extraction, (C) deep feature extraction, and (D) kernel fusion-based multi-modal analysis.
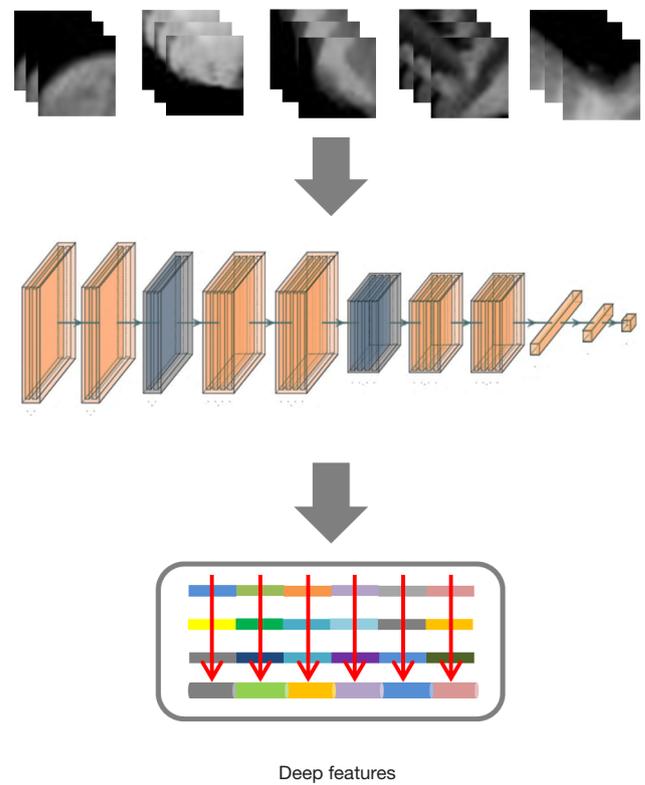
selection methods, i.e., minimum redundancy maximum correlation (mRMR) and forest-based feature selection (FF) (30,31).

Multi-modal feature integrative analysis

Since features extracted from different modalities by use of different methodologies may contain different information, simply concatenating them might not make full use of the various information to build a high-precise grading model. Therefore, a kernel fusion-based SVM (32) was used to take full advantage of these features. The kernel fusion-based SVM constructed a grading model by integrating the four feature spaces into an adaptive feature kernel space with mixed weights $\omega = \{\omega_1, \omega_2, \omega_3, \omega_4 | \omega_1 + \omega_2 + \omega_3 + \omega_4 = 1\}$, where subscripts 1, 2, 3, and 4 denote the radiomics features on T2 FLAIR and T1ce and deep features on T2 FLAIR and T1ce, respectively. Particularly, different features are usually fitted to different kernels due to the diverse data structures. The choice of the kernel type is important and depends on the data distribution and specific application. Hence, several common kernel types were compared and the best type was selected on the basis of performance on validation cohort (see Table S1). The 10-fold cross-validation was implemented to select optimal parameters of SVM during training. Finally, combining four specific feature sets, a kernel-fusion based SVM classifier was built for glioma grading and the model was evaluated on the external testing cohort.

Radiologists reading

Three radiologists with 10, 8, and 5 years of clinical experience in radiology predicted the glioma grading on the basis of the following information: unprocessed T1ce images, T2 FLAIR images, patient age, and gender. All radiologists were blinded to the reference standard, the prediction results of the proposed model, and other radiologists' predictions. The comparison between radiologists and the proposed model was performed in terms of AUC, sensitivity, and specificity.

Statistical analysis and implementation tool

All radiomics feature extraction algorithms, feature selection methods, and kernel fusion-based SVM algorithms were implemented using MATLAB 2016b (Mathworks, Natick, USA). The deep feature extraction was implemented on Python 3.6 (Python Software Foundation, Wilmington, Delaware, USA) based on the Keras package with the TensorFlow library as the backend. Descriptive demographic statistics were summarized as mean ± standard deviation, and different groups were compared using Student's t-test. AUC, sensitivity, and specificity were used to assess the performance of all models. The Youden index was used to determine the optimal sensitivity and specificity. The comparisons of AUCs were performed by Delong test. The model is available at https://github.com/zhang-de-lab/zhang-lab?from=singlemessage.

Results

Baseline characters

The grading model was developed with training cohorts (n=186) from the TCIA cohort. The rest of the TCIA cohort and the database from a local hospital were used as the internal validation cohort (n=47) and external testing cohort (n=334), respectively, to evaluate the model. We ensured that the three cohorts were independent during the study. The baseline characters of the enrolled cohorts are summarized in *Table 1*, and inclusion and exclusion criteria are defined in *Figure 2*.

Critical parameter setting of the proposed model

Tuning of CNN architecture

Since deep features learned by CNN may be affected by the architecture of CNN, we performed a sequence of experiments to validate the effectiveness of the proposed architecture, including kernel size, stride and the type of activation function and pooling (see Table S2). The best architecture (kernel size: 2×2×2, stride: 1, type of activation function: Relu, type of pooling: Average) yielded the highest AUCs of 0.81 [95% confidence interval (CI): 0.78, 0.83] and 0.82 (95% CI: 0.80, 0.84) for T2 FLAIR and T1ce, respectively, in the validation cohort.

Determination of feature dimension

For feature reduction, exhaustive experiments were conducted to select the k (from 1 to 30) discriminant features from the four obtained feature sets (i.e., T2 FLAIR radiomics, T1ce radiomics, T2 FLAIR deep, and T1ce deep features) on the basis of the average performance of the grading models on the validation cohort. As shown in *Figure 3*, k=19 achieved the optimal average AUC of 0.86

Table 1 Clinical characteristics of patients on the TCIA cohort and independent testing cohort

Characteristic	TCIA cohort			Independent testing cohort		
	GBM	LGG	P value	GBM	LGG	P value
Patients	106/233 (45.5)	127/233(54.5)		105/334 (31.4)	229/334 (68.6)	
Age (year)	58.7±13.6	46.0±13.8	<0.001*	44.3±13.3	37.5±11.6	<0.001*
Gender			0.04*			<0.001*
Woman	41/106 (38.7)	63/127 (49.6)		35/105 (33.3)	107/229 (46.7)	
Man	65/106 (61.3)	64/127 (50.4)		70/105 (66.7)	122/229 (53.3)	
Tumor grade			<0.001*			<0.001*
WHO II	0 (0.0)	63 (49.6)		0 (0.0)	112 (48.9)	
WHO III	0 (0.0)	64 (50.4)		0 (0.0)	117 (51.1)	
WHO IV	106 (100.0)	0 (0.0)		105 (100.0)	0 (0.0)	
1p19q codeletion			<0.001*			0.70
Codeletion	0/106 (0.0)	36/127 (28.3)		5/105 (4.8)	14/229 (6.1)	
Wild type	106/106 (100.0)	91/127 (71.7)		12/105 (11.4)	31/229 (13.5)	
Unknown	0/106 (0.0)	0/127 (0.0)		88/105 (83.8)	184/229 (80.4)	
IDH mutation			0.64			<0.001*
Mutation	8/106 (7.5)	103/127 (81.1)		18/105 (17.1)	110/229 (48.0)	
Wild type	98/106 (92.5)	24/127 (18.9)		46/105 (43.8)	48/229 (21.0)	
Unknown	0/106 (0.0)	0/127 (0.0)		41/105 (39.1)	71/229 (31.0)	
Histology			<0.001*			<0.001*
Astrocytoma	0/106 (0.0)	36/127 (28.3)		0/105 (0.0)	82/229 (35.8)	
Oligoastrocytoma	0/106 (0.0)	32/127 (25.2)		0/105 (0.0)	83/229 (36.2)	
Oligodendroglioma	0/106 (0.0)	59/127 (46.5)		0/105 (0.0)	44/229 (19.2)	
Glioblastoma	106/106 (100.0)	0/127 (0.0)		105/105 (100.0)	0/229 (0.0)	
Unknown	0/106 (0.0)	0/127 (0.0)		0/105 (0.0)	20/229 (8.8)	

Data in parentheses are percentages. * indicates significant difference.

among the 30 options. Finally, 19 features were selected (see Tables S3-S5).

Weight setting and kernel type selection in kernel fusion-based SVM

The grid method was used to identify the optimal assembly of four weights. The four weights with an interval of 0.05 from 0 to 1 were set and satisfied the condition of $\omega = \{\omega_1, \omega_2, \omega_3, \omega_4 | \omega_1 + \omega_2 + \omega_3 + \omega_4 = 1\}$. Among six kernel types, chi-square kernel yielded the highest AUC of 0.94 (95% CI: 0.85, 0.99) with the best coefficient combination

of 0.35, 0.15, 0.15, and 0.35 (see Table S6). Furthermore, to validate the effectiveness of kernel fusion, we compared it with direct concatenating integration. The performance of kernel fusion with AUC of 0.94 (95% CI: 0.85, 0.99) and 0.88 (95% CI: 0.84, 0.91) was superior to that of direct concatenating integration with AUC of 0.91 (95% CI: 0.79, 0.97) and 0.84 (95% CI: 0.79, 0.88) on the internal validation and external testing cohort (Delong test: P<0.05), respectively. This result indicated that kernel fusion was an effective strategy for multi-modal and multi-feature analysis.
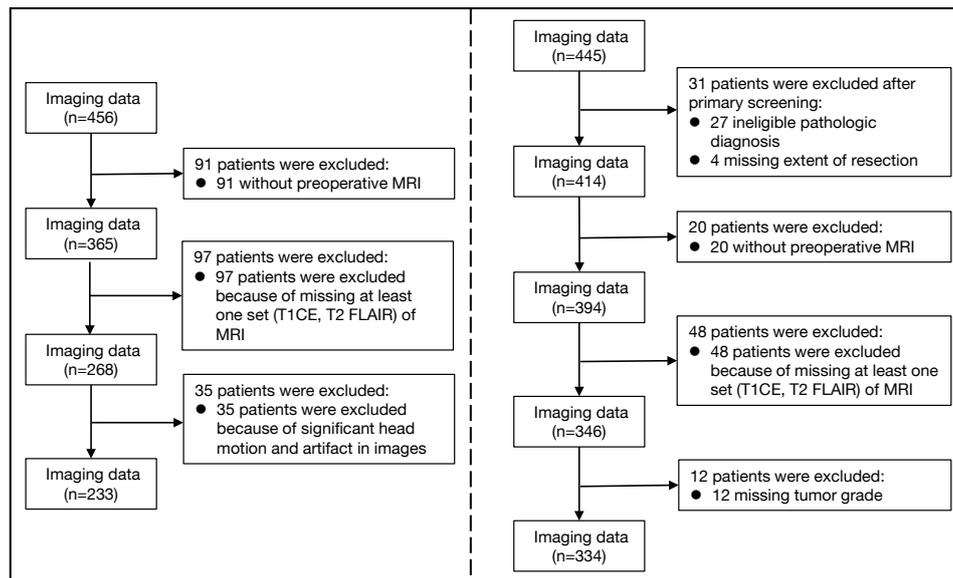
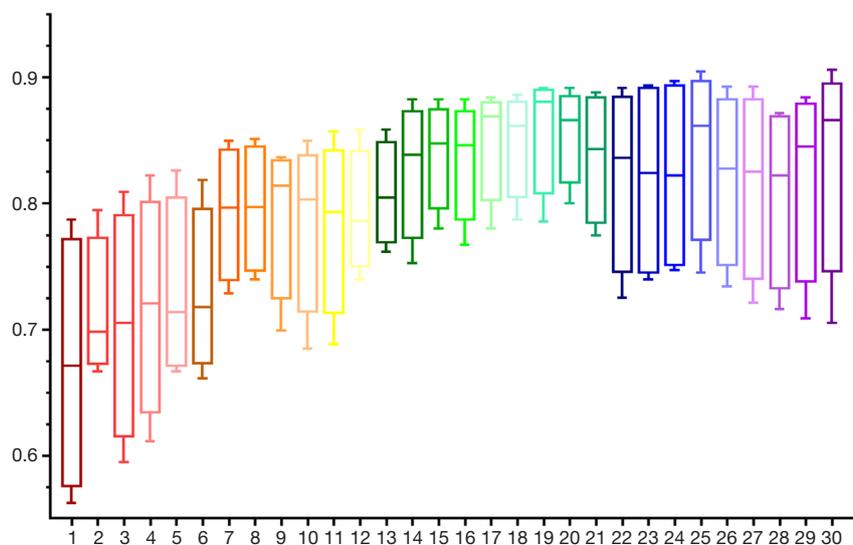**Figure 2** Patient inclusion and exclusion criteria on internal validation cohort and external testing cohort.

Figure 3 Average AUC of models on internal validation cohort with different feature dimensions from 1 to 30. The *k*=19 achieved the best performance with AUC of 0.86.

Comparison of different models

In this work, two excellent methods (i.e., radiomics and deep learning) were separately implemented on two MRI sequences (i.e., T2 FLAIR and T1ce) to obtain four specific feature sets. To validate the effectiveness of integrating radiomics and deep learning models, we compared our proposed model with these models based on an individual

approach and single MRI sequence. As shown in *Table 2*, the best grading performance was obtained by the proposed integrative model with the AUC of 0.94 (95% CI: 0.85, 0.99) on internal validation cohort and 0.88 (95% CI: 0.84, 0.91) on external testing cohort. The performances were inferior when only a single method with a single MRI sequence was used. For internal validation cohort, the AUC was 0.88 (95% CI: 0.75, 0.95) for radiomics on T2 FLAIR,

Page 8 of 12

Ning et al. Gliomas grading using radiomics and deep features

Table 2 Comparison of the proposed method with the models based on different methodologies and modalities

Method	Validation AUC	Sensitivity (%)	Specificity (%)	Testing AUC	Sensitivity (%)	Specificity (%)
R_T2 FLAIR	0.88 (0.75, 0.95)	76 (16/21)	85 (22/26)	0.81 (0.77, 0.85)	80 (84/105)	71 (163/229)
R_T1ce	0.87 (0.74, 0.95)	81 (17/21)	77 (20/26)	0.80 (0.75, 0.84)	71 (75/105)	79 (180/229)
R_T2 FLAIR + T1ce	0.92 (0.80, 0.98)	86 (18/21)	85 (22/26)	0.85 (0.81, 0.89)	85 (89/105)	76 (175/229)
D_T2 FLAIR	0.86 (0.73, 0.95)	81 (17/21)	77 (20/26)	0.79 (0.75, 0.84)	75 (79/105)	76 (175/229)
D_T1ce	0.88 (0.75, 0.95)	71 (15/21)	88 (23/26)	0.80 (0.76, 0.84)	80 (84/105)	69 (158/229)
D_T2 FLAIR + T1ce	0.91 (0.79, 0.97)	81 (17/21)	88 (23/26)	0.84 (0.80, 0.88)	82 (86/105)	77 (176/229)
R + D_ T2 FLAIR	0.90 (0.77, 0.97)	81 (17/21)	85 (22/26)	0.82 (0.78, 0.86)	78 (82/105)	76 (173/229)
R + D_T1ce	0.90 (0.80, 0.98)	86 (18/21)	81 (21/26)	0.82 (0.77, 0.86)	81 (85/105)	75 (172/229)
Proposed method	0.94 (0.85, 0.99)	86 (18/21)	92 (24/26)	0.88 (0.85, 0.92)	88 (92/105)	81 (186/229)

Note. R_: Radiomics, D_: Deep learning, R + D_: Radiomics + Deep learning

Table 3 Performance of models using different feature selection methods and classifiers

Method	Validation AUC	Sensitivity (%)	Specificity (%)	Testing AUC	Sensitivity (%)	Specificity (%)
mRMR + SVM	0.92 (0.80, 0.98)	86 (18/21)	85 (22/26)	0.85 (0.81, 0.89)	87 (91/105)	74 (169/229)
FF + SVM	0.90 (0.77, 0.97)	86 (18/21)	88 (23/26)	0.84 (0.80, 0.88)	83 (87/105)	75 (171/229)
Relief + RF	0.92 (0.80, 0.98)	81 (17/21)	92 (24/26)	0.86 (0.82, 0.90)	85 (89/105)	81 (186/229)
Relief + SVM	0.94 (0.85, 0.99)	86 (18/21)	92 (24/26)	0.88 (0.84, 0.91)	88 (92/105)	81 (186/229)

0.87 (95% CI: 0.74, 0.95) for radiomics on T1ce, 0.86 (95% CI: 0.73, 0.95) for deep learning on T2 FLAIR, and 0.88 (95% CI: 0.75, 0.95) for deep learning on T1ce. For external testing cohort, the AUC was 0.81 (95% CI: 0.77, 0.85), 0.80 (95% CI: 0.75, 0.84), 0.79 (95% CI: 0.75, 0.84), and 0.80 (95% CI: 0.76, 0.84), respectively. By contrast, the results were improved whether conducting each single approach on multi-modal images or combining two approaches on a single sequence. The AUC for radiomics on multi-modal images, deep learning on multi-modal images, combined methods on T2 FLAIR, and combined methods on T1ce was 0.92 (95% CI: 0.80, 0.98), 0.91 (95% CI: 0.79, 0.97), 0.90 (95% CI: 0.77, 0.97), and 0.90 (95% CI: 0.80, 0.98) for internal validation cohort; and 0.85 (95% CI: 0.81, 0.89), 0.84 (95% CI: 0.80, 0.88), 0.82 (95% CI: 0.78, 0.86), and 0.82 (95% CI: 0.78, 0.86) for external testing cohort, respectively.

Some conclusions could be drawn: (I) the combination of radiomics and deep learning method achieved the better performance compared with individual radiomics or deep learning method whether single or multiple modal MR images were used; (II) the models based on multiple

modal MR images were superior to these models based on individual T1ce or T2 FLAIR MR images. In addition, the comparison results of the proposed integrative model with other models with different feature selection methods and classifiers are listed in *Table 3*, from which we can see that relief algorithm and SVM showed the better results when compared with other methods (Delong test: P<0.05).

Radiologists reading

Table 4 shows the sensitivity and specificity of three radiologists and the proposed model. For internal validation and external testing cohorts performed by the proposed model, the sensitivity at the optimal threshold of the Youden index was 86% and 88%, respectively, while the specificity was 92% and 81%, respectively. The sensitivity of the radiologists ranged between 80% and 90%, while the specificity was between 72% and 85%. *Figure 4A,B* show the ROC of our proposed model on internal validation and external testing cohort, respectively. For comparison, the points representing the sensitivity and specificity of the three radiologists for grading glioma are also shown

Table 4 Comparison between three radiologists and the proposed method for gliomas grading

Radiologist	Validation cohort		Testing cohort	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Reader 1	90 (19/21)	77 (20/26)	84 (88/105)	75 (172/229)
Reader 2	86 (18/21)	85 (22/26)	81 (85/105)	78 (178/229)
Reader 3	86 (18/21)	73 (19/26)	80 (84/105)	72 (165/229)
Proposed method	86 (18/21)	92 (24/26)	88 (92/105)	81 (186/229)
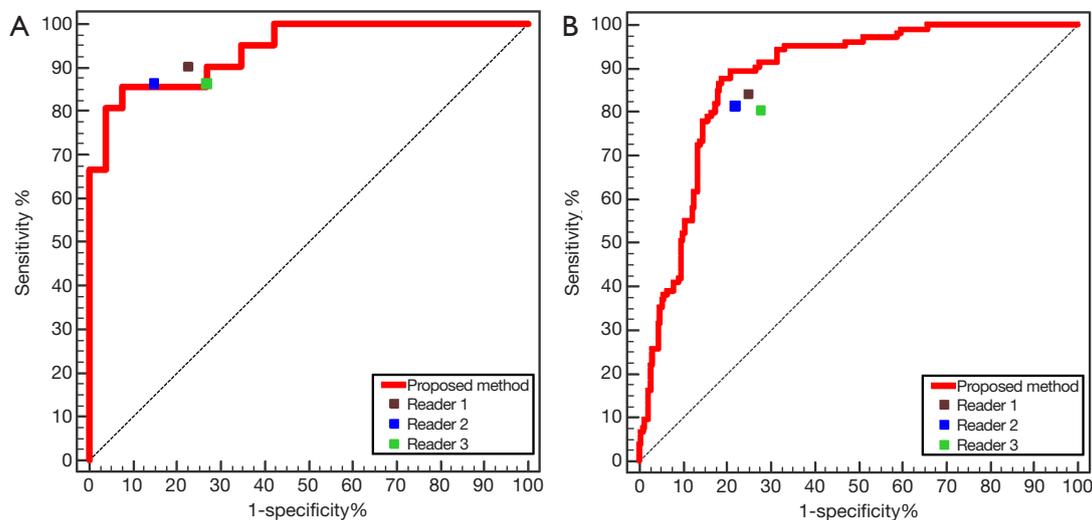
Figure 4 ROC curve of the proposed method on internal validation cohort (A) and external testing cohort (B). Three points representing sensitivity and specificity of three readers were plotted.

in *Figure 4*. These points were close to the ROC of the proposed model.

Discussion

In this study, we presented a multi-modal MRI-based grading analysis by combining the radiomics and deep learning technologies. The proposed grading model took full advantage of the global radiomics and local deep features from multi-modal MR images and achieved outstanding results in the internal validation and external testing cohorts. This finding demonstrated the feasibility of integrating global radiomics and local deep features to develop a model for glioma grading.

Most previous studies only focused on individual radiomics or deep learning approaches to conduct a quantitative analysis of glioma grading (5,14,33). However, few researches explored whether combining radiomics and

deep learning approaches can improve the performance of a grading model for gliomas. According to the comparison results between the proposed method and two individual approaches based on single and multiple modal MR images, it could be found that (I) the integration of radiomics and deep learning model outperformed any single method whether single or multiple modal MR images were used; (II) multi-modal MR images could provide more information than single modal MR images regardless of the methodologies. Meanwhile, the sensitivity and specificity of the proposed model were comparable to those of clinical radiologists, which further emphasized the promising preliminary results of the proposed method for glioma grading. In addition, for radiologists, the sensitivity on the testing cohort were lower than that of the validation cohort, which might be caused by more testing samples, and affected by the experience, and some subjective factors (such as fatigue and attention). Moreover, the proposed

model could provide robust sensitivity in glioma grading. Meanwhile, the lower specificity of the proposed model on the testing cohort needs to be improved, as the high number of false-positive gliomas would require the clinical radiologists to verify the actual presence of disease, which would increase the overall time of image interpretation. A potential reason of lower specificity was that cohorts used in our studies were from different centers, in which different fields of view, spatial resolutions, section thickness, and intersection gaps of the different sequences were acquired during the MRI examination.

One of the challenges of this study is the high feature dimension, which is inclined to trap in overfitting; and the modeling with high-dimension features is also time-consuming. In our work, a total of 10,324 radiomics and 13,824 deep features were extracted for singular modality. Therefore, the following three strategies were used to select discriminant features and suppress overfitting in our work. (I) The proposed CNN was equipped with pooling and dropout operators, which prevented the redundancy of features and improved robustness. (II) Patch pooling was used to integrate features from the patch level to the patient level rather than directly connecting all deep features. This procedure could also address the number of inconsistency caused by the patch strategy. (III) The relief algorithm was conducted on four feature sets, and the reduced dimension remained the same to fit the subsequent analysis. The second challenge is taking full advantage of multi-modal information and multiple types of features. A kernel fusion-based SVM was introduced to weigh different features instead of direct concatenation. In this way, features extracted from different modal MRIs by different technologies were effectively integrated, and the classification results also demonstrated the fusion strategy was helpful to analysis of multi-modal features.

Our study has several deficiencies. First, tumor segmentation was still a manual process, which was time-consuming and depended on the experience of the radiologists. Second, only T1ce and T2 FALIR sequences were used in our study; however, more than two modal MRI data can be collected for further analysis. Lastly, this work was a retrospective study, and a prospective cohort is required to further evaluate the performance of the glioma grading model.

In conclusion, an effective integrative strategy that combined two popular technologies, i.e., radiomics and deep learning, was proposed for grading gliomas. The approach adopted the kernel fusion method to build a discriminative SVM classifier based on postcontrast enhanced T1-weighted and T2 fluid-attenuated inversion recovery sequences. Furthermore, an independent external testing cohort was used to assess the generalization performance of the proposed grading model. The promising results demonstrated the feasibility of integrating radiomics and deep learning based on multi-modal magnetic resonance images for grading gliomas.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at http://dx.doi.org/10.21037/atm-20-4076

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-20-4076). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This retrospective study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Institutional Review Board of Nanfang Hospital (Guangzhou, Guangdong, China; ID: NFEC-2020-251) and individual consent for this retrospective study was waived.

original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Ostrom QT, Gittleman H, Liao P, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007-2011. Neuro Oncol 2014;16:iv1-iv63.

2. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. Acta Neuropathol 2016;131:803-20.

3. Arvinda HR, Kesavadas C, Sarma PS, et al. Glioma grading: sensitivity, specificity, positive and negative predictive values of diffusion and perfusion imaging. J Neurooncol 2009;94:87-96.

4. Whittle IR. The dilemma of low grade glioma. J Neurol Neurosurg Psychiatry 2004;75:ii31-6.

5. Tian Q, Yan LF, Zhang X, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI: Radiomics Approach for Glioma Grading. J Magn Reson Imaging 2018;48:1518-28.

6. Jackson RJ, Fuller GN, Abi-Said D, et al. Limitations of stereotactic biopsy in the initial management of gliomas. Neuro Oncol 2001;3:193-200.

7. Glantz MJ, Burger PC, Herndon JE, et al. Influence of the type of surgery on the histologic diagnosis in patients with anaplastic gliomas. Neurology 1991;41:1741-4.

8. Field M, Witham TF, Flickinger JC, et al. Comprehensive assessment of hemorrhage risks and outcomes after stereotactic brain biopsy. J Neurosurg 2001;94:545-51.

9. Huang YQ. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. J Clin Oncol 2016;34:2157-64.

10. Kim M, Jung SY, Park JE, et al. Diffusion- and perfusion-weighted MRI radiomics model may predict isocitrate dehydrogenase (IDH) mutation and tumor aggressiveness in diffuse lower grade glioma. Eur Radiol 2020;30:2142-51.

11. Yang X, He J, Wang J, et al. CT-based radiomics signature for differentiating solitary granulomatous nodules from solid lung adenocarcinoma. Lung Cancer 2018;125:109-14.

12. Hsieh KL, Lo CM, Hsiao CJ. Computer-aided grading of gliomas based on local and global MRI features. Comput Methods Programs Biomed 2017;139:31-8.

13. Lu CF, Hsu FT, Hsieh LC, et al. Machine Learning-Based Radiomics for Molecular Subtyping of Gliomas. Clin Cancer Res 2018;24:4429-36.

14. Su C, Jiang J, Zhang S, et al. Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour. Eur Radiol 2019;29:1986-96.

15. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017;318:2199-210.

16. Chen PJ, Lin MC, Lai MJ, et al. Accurate classification of diminutive colorectal polyps using computer aided analysis. Gastroenterology 2018;154:568-75.

17. Wang K, Lu X, Zhou H, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut 2019;68:729-41.

18. Decuyper M, Bonte S, Van Holen R. Binary Glioma Grading: Radiomics versus Pre-trained CNN Features. In: Frangi A, Schnabel J, Davatzikos C, et al. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Cham: Springer, 2018:498-505.

19. Huang P, Li D, Jiao Z, et al. CoCa-GAN: Common-Feature-Learning-Based Context-Aware Generative Adversarial Network for Glioma Grading. In: Shen Dm Liu T, Peters TM, et al. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Cham: Springer, 2019.

20. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.

21. Chaudhary K, Poirion OB, Lu L, et al. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. Clin Cancer Res 2018;24:1248-59.

22. Wang S, Zhou M, Liu Z, et al. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. Med Image Anal 2017;40:172-83.

23. Sajjad M, Khan S, Muhammad K, et al. Multi-Grade Brain Tumor Classification using Deep CNN with Extensive Data Augmentation. J Comput Sci 2018;30:174-82.

24. Banerjee S, Mitra S, Masulli F, et al. Deep Radiomics for Brain Tumor Detection and Classification from Multi-Sequence MRI. arXiv:1903.09240v1 [preprint]. 2019 [cited 2019 Mar 21]. Available online: https://arxiv.org/abs/1903.09240

25. Decuyper M, Holen VR. Fully Automatic Binary Glioma Grading based on Pre-Therapy MRI using 3D

Page 12 of 12

Ning et al. Gliomas grading using radiomics and deep features

Convolutional Neural Networks. arXiv:1908.01506 [preprint]. 2019 [cited 2019 Aug 5]. Available online: https://arxiv.org/abs/1908.01506v1

26. Vallières M, Freeman CR, Skamene S, et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol 2015;60:5471-96.

27. Tieleman T, Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA 2012:4:26-31.

28. Ning Z, Luo J, Li Y, et al. Pattern Classification for Gastrointestinal Stromal Tumors by Integration of Radiomics and Deep Convolutional Features. IEEE J Biomed Health Inform 2019;23:1181-91.

29. Kira K, Rendell LA. The feature selection problem: traditional methods and a new algorithm. Available online: https://www.aaai.org/Papers/AAAI/1992/AAAI92-020.pdf

30. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27:1226-38.

31. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007;23:2507-17.

32. Luo J, Ning Z, Zhang S, et al. Bag of deep features for preoperative prediction of sentinel lymph node metastasis in breast cancer. Phys Med Biol 2018;63:245014.

33. Yang Y, Yan LF, Zhang X, et al. Glioma Grading on Conventional MR Images: A Deep Learning Study With Transfer Learning. Front Neurosci 2018;12:804.

Role of thoracic radiotherapy in extensive stage small cell lung cancer: a systemic review and meta-analysis

Ao-Mei Li[1#], Han Zhou[1#], Yang-Yang Xu[2#], Xiao-Qin Ji[1], Tian-Cong Wu[1], Xi Yuan[1], Chang-Chen Jiang[1], Xi-Xu Zhu[1], Ping Zhan[2,3], Ze-Tian Shen[1]

[1]Department of Radiation Oncology, Jinling Hospital, Medical School of Nanjing University, Nanjing, China; [2]Department of Respiratory and Critical Care Medicine, Jinling Hospital, Nanjing Medical University, Nanjing, China; [3]Department of Respiratory and Critical Care Medicine, Jinling Hospital, Nanjing University School of Medicine, Nanjing, China

Contributions: (I) Conception and design: P Zhan, ZT Shen; (II) Administrative support: CC Jiang, P Zhan; (III) Provision of study materials or patients: XQ Ji, TC Wu, X Yuan; (IV) Collection and assembly of data: AM Li, H Zhou, YY Xu; (V) Data analysis and interpretation: AM Li, H Zhou, YY Xu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.
[#]These authors contributed equally to this work.

Correspondence to: Ping Zhan. Department of Respiratory and Critical Care Medicine, Jinling Hospital, Nanjing University School of Medicine, #305, East Zhongshan Road, Nanjing 210002, China. Email: zhanping207@163.com; Ze-Tian Shen. Department of Radiation Oncology, Jinling Hospital, Medical School of Nanjing University, Nanjing, Jiangsu, 210002, China. Email: shen-zetian@163.com.

Background: The role of thoracic consolidation radiotherapy in patients with extensive stage small cell lung cancer (ES-SCLC) remains controversial. This study aimed to evaluate the efficacy of thoracic radiotherapy (TRT) in these patients.

Methods: A systematic literature search was performed in PubMed, Embase, and the Cochrane library to identify qualified clinical studies. The hazard ratios (HRs) and 95% confidence intervals (CIs) of overall survival (OS), progression-free survival (PFS) and local recurrence-free survival (LRFS) were extracted, and toxicity of the TRT group versus non-TRT group was analyzed.

Results: A total of 12 studies were included in this meta-analysis, including 936 patients in the TRT group and 1,059 patients in the non-TRT group. The combined results showed that TRT significantly improved OS (HR =0.65; 95% CI: 0.55–0.77, P<0.00001), PFS (HR =0.64; 95% CI: 0.56–0.72, P<0.00001) and LRFS (HR =0.38, 95% CI: 0.26–0.53, P<0.00001). Subgroup analysis showed that OS benefits were observed in patients receiving sequential TRT (HR =0.67; 95% CI: 0.54–0.84, P=0.0006). The addition of TRT significantly improved OS in patients over 65 years of age (HR =0.55; 95% CI: 0.40–0.74, P=0.0001). For patients with only one organ metastasis, there was no significant difference in OS between the two groups (HR =0.61; 95% CI: 0.36–1.01, P=0.06). There was no statistical difference in hematologic toxicity (leukopenia, thrombocytopenia, anemia) and non-hematologic toxicity (nausea or vomiting) between the two groups. The incidence of grade ≥3 esophageal toxicity was 4.6% in the TRT group and 0% in the non-TRT group (P=0.0001). Grade ≥3 bronchopulmonary toxicity was 2.9% in the TRT group and 0.8% in the non-TRT group (P=0.02).

Conclusions: TRT improves OS, PFS and LRFS in patients with ES-SCLC, with a low increase in esophageal and bronchopulmonary toxicity. More randomized controlled trials (RCTs) are expected to confirm our conclusions.

PROSPERO registration number: CRD42020190575.

Keywords: Small cell lung cancer (SCLC); thoracic radiotherapy (TRT); meta-analysis

Page 2 of 13

Li et al. Role of TRT in ES-SCLC

Introduction

Small cell lung cancer (SCLC) accounts for 10–15% of all lung cancers. About 70% of the patients are already in the extensive stage at the initial diagnosis, with a median overall survival (OS) of only 8–10 months (1,2). So far, platinum-based chemotherapy is still the standard treatment for extensive stage small cell lung cancer (ES-SCLC), with a response rate of up to 60–80%. However, most patients (approximately 90%) develop intrathoracic progression within 1 year after completion of chemotherapy. There are few effective treatment options, and the 2-year survival rate is only 5% (3,4).

Since SCLC is sensitive to radiotherapy, many studies have explored the effect of consolidation thoracic radiotherapy (TRT) in ES-SCLC. A randomized controlled trial (RCT) published in 1999 was the first to suggest that the addition of TRT improved OS and first relapse-free survival in ES-SCLC. Patients with complete response (CR) at a distant site and at least partial response (PR) at a local site were selected in the study as they had a better prognosis (5). A subsequent RCT demonstrated TRT improved long-term survival in ES-SCLC, with the 2-year OS rate of 13%. In addition, TRT reduced the incidence of intrathoracic progression ($P<0.0001$) (6). Recent retrospective analyses and database-based studies have also demonstrated the survival benefits of thoracic consolidation radiotherapy (7-9). However, the latest randomized trial RTOG-0937 found that despite improved progression-free survival (PFS), the median survival in the TRT group (13.8 months) was worse than that in the prophylactic cranial irradiation (PCI) group (15.8 months) (10). In another retrospective study, no PFS or OS benefit of TRT was observed in patients with brain/liver/multi-metastasis (11). Therefore, the role of thoracic consolidation radiotherapy in patients with ES-SCLC remains controversial.

In this meta-analysis, we compared the effects of thoracic consolidation radiotherapy versus chemotherapy alone on survival of patients with ES-SCLC by systematically reviewing RCTs and retrospective studies.

We present the following article in accordance with the PRISMA reporting checklist (available at http://dx.doi.org/10.21037/atm-20-5765).

Methods

Search strategy

The review was registered in PROSPERO with the registration number CRD42020190575. PubMed, Embase, and the Cochrane library were systematically searched without any language restriction. All electronic database resources are available from the Nanjing Medical University Library. Our team has mastered the search strategy of electronic databases under the guidance of a medical librarian. Search terms were: ("extensive" or "advanced" or "metastatic" or "stage IV") and ("radiation" or "radiotherapy") and ("small cell lung cancer" or "small cell lung carcinoma" or "SCLC"). The last search deadline was July 13, 2020. When necessary, the references quoted in the article were manually retrieved.

Inclusion and exclusion criteria

The studies which compared the clinical efficacy of TRT combined with chemotherapy and chemotherapy alone for ES-SCLC were eligible for inclusion. Studies that met the following criteria were excluded: (I) insufficient survival data; (II) not ES-SCLC; (III) review, comments, survey, conference abstract, meta-analysis; (IV) not platinum-based chemotherapy as a control group; (V) not TRT.

Data extraction

Two authors (AML and YYX) independently screened eligible studies and extracted the data. Disagreements were resolved by consulting the third author (HZ). The following information was extracted: author, year of publication, country, study period, number of participants, study type, treatment regimen (including chemotherapy, TRT and PCI), and survival data. We evaluated OS, PFS, local recurrence-free survival (LRFS) and toxicity in the meta-analysis. If both univariate and multivariate analyses of survival data were provided in the study, multivariate analysis results were applied because they reduced confounding bias. Similarly, if survival data both before and after propensity score matching (PSM) were presented in the article, the results after PSM were used.

Quality assessment and statistical analysis

The quality of retrospective studies was evaluated using the Newcastle-Ottawa Scale (NOS). The possible NOS score of each article ranged from 0 to 9, and studies with a score greater than 6 were considered to be of high quality (12). The quality of RCTs was assessed using the Cochrane risk of bias tool (13). Two authors (AML and HZ) independently
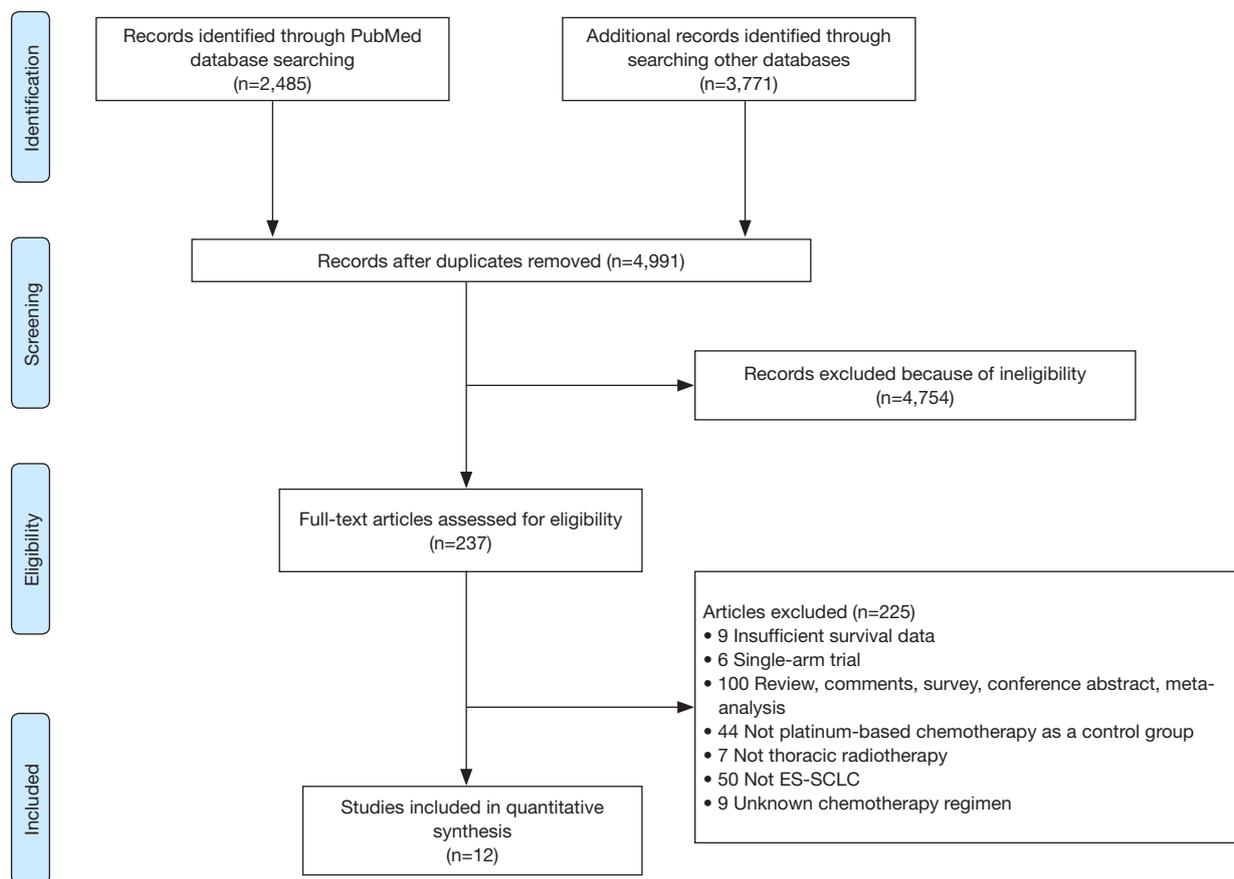
Figure 1 Flow diagram of study selection procedure.

evaluated the included studies. Any inconsistency was resolved by consensus.

Review Manager version 5.3 (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration) was used for statistical analysis. A P<0.05 was considered statistically significant. The effects of thoracic consolidation radiation on OS, PFS and LRFS in patients with ES-SCLC were evaluated by hazard ratios (HRs) and the corresponding 95% confidence interval (CIs). If the HRs and 95% CIs were not directly reported in the included article, we used Engauge Digitizer version 11.1 (http://sourceforge.net/projects/digitizer/) to estimate them from the Kaplan-Meier curve (14,15). Risk ratio (RR) was used to analyze toxicity. Heterogeneity was tested by Cochrane's Q test and I-square (I^2). The random-effects model was used when significant heterogeneity was observed (P<0.1 or I^2>50%); otherwise, the fixed-effects model was applied. Potential publication bias was assessed by funnel plots. Sensitivity analysis was performed when necessary to explore possible sources of

heterogeneity.

Results

Study selection and characteristics

We initially retrieved 6,256 studies from the electronic databases, of which 1,265 were duplicates. A total of 4,754 studies were excluded by screening titles and abstracts. After reading the full text of the remaining 237 studies, 12 studies were eventually included in the meta-analysis: 3 RCTs and 9 retrospective studies (5-6,8-11,16-21). The selection process of the study is shown in *Figure 1*.

A total of 1,995 patients with ES-SCLC were included in this meta-analysis, of whom 936 received thoracic consolidation radiotherapy (TRT group) and 1,059 received chemotherapy alone (non-TRT group). The basic characteristics of the included studies are shown in *Table 1*. Four retrospective studies performed PSM (8,16-18). Two retrospective studies evaluated the role of TRT in

Page 4 of 13

Li et al. Role of TRT in ES-SCLC

Table 1 Baseline characteristics of included studies

Source	Country	Study period	No. of patients	Study type	ChT regimen	TRT approach	TRT regimen	TRT Y/N	PCI approach	PCI regimen	PCI Y/N	Outcome
Zhu, 2011	China	January 2003–December 2006	119	Retrospective cohort	EP/EC	Sequential (after ChT)	40–60 Gy at 1.8–2.0 Gy/Fr	60/59	NA	30 Gy	2/117	OS, PFS and Toxicity
Gore, 2017	US, Canada, Israel	March 2010–February 2015	86	Randomized controlled (phase-II)	Platinum based chemotherapy	Sequential (after ChT)	45 Gy/15 Fr	44/42	Concurrent with TRT	25 Gy/10 Fr	86/0	OS, Time to progression and Toxicity
Jeremic, 1999	Yugoslavia	January 1988–June 1993	109	Randomized controlled (phase-III)	EP/EC	ChT followed by concurrent TRT-ChT	54 Gy/36 Fr	55/54	Sequential (after TRT)	25 Gy/10 Fr	109/0	OS, LRFS, DMFS and PFS
Slotman, 2015	Netherlands, UK, Norway, Belgium	February 2009–December 2012	495	Randomized controlled (phase-III)	Standard chemotherapy (platinum etoposide)	Sequential (after ChT)	30 Gy/10 Fr	247/248	Concurrent with TRT	20 Gy/5 Fr, 25 Gy/10 Fr, 30 Gy/10 Fr, 30 Gy/12 Fr, 30 Gy/15 Fr	495/0	OS, PFS and Toxicity
Qi, 2019	China	January 2010–January 2018	80, aged ≥65 years	Match-controlled	EP/EC	Sequential (after ChT)/ Concurrent with ChT	30–60 Gy at 1.8–3 Gy/Fr	40/40	NA	NA	NA	OS, PFS, LRFS
An, 2017	China	January 2006–December 2013	118, aged ≥65 years	Retrospective cohort	EP/EC	NA	48–60 Gy at 1.8–2.0 Gy/Fr	59/59	Sequential (after TRT)	25 Gy /10 Fr	43/49	OS, PFS and Toxicity
Deng, 2019	China	January 2007–December 2012	144	Match-controlled	EP/EC	Sequential (after ChT)	32–67 Gy/25–33 Fr	72/72	NA	NA	0/144	OS, PFS
Luan, 2015	China	January 2005–December 2008	165	Retrospective cohort	EP/EC	Sequential (after ChT)	40–62 Gy at 1.5–2 Gy/Fr	82/83	NA	30 Gy	5/160	OS, PFS and Toxicity
Luo, 2017	China	January 2011–December 2015	90	Match-controlled	EP/EC	Sequential (after ChT)	45–60 Gy/15–30 Fr	45/45	Sequential (after TRT)	25 Gy/10 Fr	3/87	OS, PFS, LRFS

Table 1 (*continued*)

Table 1 (*continued*)

Source	Country	Study period	No. of patients	Study type	ChT regimen	TRT approach	TRT regimen	TRT Y/N	PCI approach	PCI regimen	PCI Y/N	Outcome
Qin, 2016	China	January 2000–March 2010	94	Retrospective cohort	EC/EP or sequential paclitaxel with cisplatin	Sequential (after ChT)/ Concurrent with ChT	40–60 Gy at 1.8–2.0 Gy/Fr	32/62	NA	NA	5/89	OS, PFS and Toxicity
Zhang, 2019	China	January2005–January 2016	305	Retrospective cohort	EC/EP or platinum-based chemotherapy	Sequential (after ChT)	40–60 Gy/20–30 Fr or 30–45 Gy/10–15 Fr	105/200	NA	25 Gy	34/271	OS, PFS and Toxicity
Xu, 2017	China	May 2010–May 2015	190	Match-controlled	EC/EP	Sequential (after ChT)	30–60 Gy at 1.8–3 Gy/Fr	95/95	NA	NA	11/179	OS, PFS, LC

ChT, chemotherapy; TRT, thoracic radiotherapy; PCI, prophylactic cranial irradiation; Y, yes; N, no; EP, etoposide and cisplatin; EC, etoposide and carboplatin; OS, overall survival; PFS, progression free survival; NA, not available; Gy, gray; Fr, fractions; LC, local control; DMFS, distant metastasis-free survival; LRFS, local recurrence-free survival.

patients aged ≥65 years (16,19). One retrospective study analyzed the prognosis of patients with oligometastases and polymetastases separately (18).

Quality assessment

All the retrospective studies scored ≥7, and were considered to have high quality. As shown in *Table 2*, all three RCTs claimed randomization, but no methods for random assignment sequence generation and allocation concealment were reported. The age distribution of patients in the experimental group and the control group in trial RTOG-0937 was unbalanced (10).

OS

All twelve studies reported OS results. Combined analysis showed that TRT significantly improved OS (HR =0.65; 95% CI: 0.55–0.77, P<0.00001) (*Figure 2A*). As significant heterogeneity was observed (I²=62%), the random-effects model was applied.

PFS

All twelve studies reported PFS results. The pooled HR with a random effects model was 0.64 (95% CI: 0.56–0.72, P<0.00001) (*Figure 2B*), which indicated that the TRT group had significantly better PFS than non-TRT group.

LRFS

Four studies compared LRFS between TRT and non-TRT groups. The combined analysis with a random effects model showed that TRT significantly improved LRFS (HR =0.38; 95% CI: 0.26–0.53, P<0.00001) (*Figure 2C*).

Subgroup analysis

We performed subgroup analysis based on study type to assess the effect of TRT on OS and PFS. The combined HRs for OS of three RCTs and nine retrospective studies were 0.88 (95% CI: 0.65–1.18, P=0.39) and 0.59 (95% CI: 0.49–0.71, P<0.00001), respectively (*Figure 3A*). The random-effects model was applied as significant heterogeneity was observed (I²=54% for RCTs). The pooled HR for PFS of three RCTs was 0.71 (95% CI: 0.61–0.82, P<0.00001), and it was 0.61 (95% CI: 0.52–0.73, P<0.00001) for retrospective studies (*Figure 3B*). The random-effects

Page 6 of 13

Li et al. Role of TRT in ES-SCLC

Table 2 Quality assessment of included studies

Source	Sequence generation	Allocation concealment	Blinding	Incomplete outcome data	Selective outcome reporting	Other bias
Gore, 2017	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	High risk
Jeremic, 1999	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Low risk
Slotman, 2015	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Low risk

	Selection	Comparability	Outcome	Score
Qi, 2019	★★★	★★	★★	7
An, 2017	★★★	★★	★★★	8
Deng, 2019	★★★★	★★	★★★	9
Luan, 2015	★★★★	★	★★	7
Luo, 2017	★★★★	★★	★★	8
Qin, 2016	★★★	★	★★★	7
Zhang, 2019	★★★★	★	★★	7
Xu, 2017	★★★★	★★	★	7
Zhu, 2011	★★★★	★	★★★	8

★, one point. Randomized trials were assessed by the Cochrane risk of bias tool. Retrospective studies were assessed by NOS method. A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability, according to the instruction of NOS.

model was applied as I² of the pooled retrospective studies was 57%.

Comparison of OS between the TRT and non-TRT groups according to clinicopathologic features

Nine studies reported OS in patients receiving sequential TRT in combination with platinum-based chemotherapy. The pooled analysis using a random effects model showed that the sequential TRT group had better OS than the non-TRT group (HR =0.67; 95% CI: 0.54–0.84, P=0.0006) (*Figure 4A*). For patients aged ≥65 years, the TRT group had better OS than the non-TRT group (HR =0.55; 95% CI: 0.40–0.74, P=0.0001). The fixed effects model was used as I²=0% (*Figure 4B*). Three studies reported OS in patients with oligometastatic disease (defined as only one organ metastasis). The pooled HR with a random effects model was 0.61 (95% CI: 0.36–1.01, P=0.06) (*Figure 4C*).

Toxicity

Grade III or higher toxicity was analyzed in 6 studies,

including 570 cases in the TRT group and 694 cases in the non-TRT group. The pooled results with a fixed effects model indicated the TRT group had a higher risk of grade III or higher toxicity than the non-TRT group (RR =1.34; 95% CI: 1.17–1.53, P<0.0001) (*Figure 5*). *Table 3* shows the comparison of common grade III or higher toxicity between the TRT group and the non-TRT group. The pooled RRs were 1.26 (95% CI: 0.82–1.95, P=0.29) for grade ≥3 leucopenia, 0.67 (95% CI: 0.44–1.03, P=0.07) for grade ≥3 thrombocytopenia, 0.92 (95% CI: 0.53–1.59, P=0.76) for grade ≥3 anemia, 0.61 (95% CI: 0.35–1.08, P=0.09) for grade ≥3 nausea or vomiting, 13.89 (95% CI: 3.63–53.19, P=0.0001) for grade ≥3 esophageal toxicity and 2.63 (95% CI: 1.21–5.72, P=0.02) for grade ≥3 bronchopulmonary toxicity.

Sensitivity analysis and publication bias

The funnel plot was applied to test publication bias. No substantial publication bias was observed with respect to OS (*Figure 6*). We also performed sensitivity analysis by deleting one study at a time. The results showed that the effect of TRT on OS, PFS and LRFS was reliable.
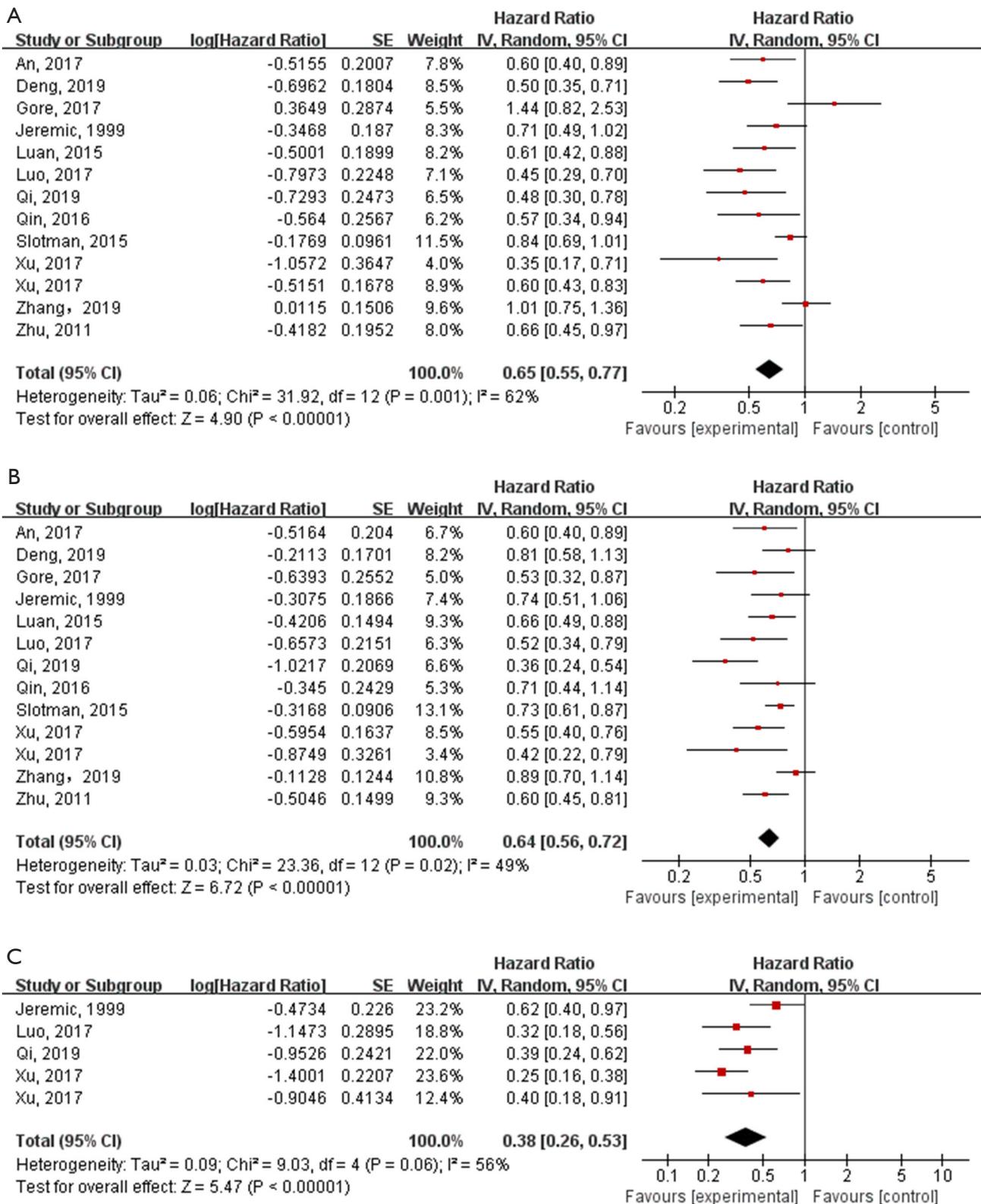
A

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
An, 2017	-0.5155	0.2007	7.8%	0.60 [0.40, 0.89]
Deng, 2019	-0.6962	0.1804	8.5%	0.50 [0.35, 0.71]
Gore, 2017	0.3649	0.2874	5.5%	1.44 [0.82, 2.53]
Jeremic, 1999	-0.3468	0.187	8.3%	0.71 [0.49, 1.02]
Luan, 2015	-0.5001	0.1899	8.2%	0.61 [0.42, 0.88]
Luo, 2017	-0.7973	0.2248	7.1%	0.45 [0.29, 0.70]
Qi, 2019	-0.7293	0.2473	6.5%	0.48 [0.30, 0.78]
Qin, 2016	-0.564	0.2567	6.2%	0.57 [0.34, 0.94]
Slotman, 2015	-0.1769	0.0961	11.5%	0.84 [0.69, 1.01]
Xu, 2017	-1.0572	0.3647	4.0%	0.35 [0.17, 0.71]
Xu, 2017	-0.5151	0.1678	8.9%	0.60 [0.43, 0.83]
Zhang, 2019	0.0115	0.1506	9.6%	1.01 [0.75, 1.36]
Zhu, 2011	-0.4182	0.1952	8.0%	0.66 [0.45, 0.97]
Total (95% CI)			**100.0%**	**0.65 [0.55, 0.77]**

Heterogeneity: Tau² = 0.06; Chi² = 31.92, df = 12 (P = 0.001); I² = 62%
Test for overall effect: Z = 4.90 (P < 0.00001)

B

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
An, 2017	-0.5164	0.204	6.7%	0.60 [0.40, 0.89]
Deng, 2019	-0.2113	0.1701	8.2%	0.81 [0.58, 1.13]
Gore, 2017	-0.6393	0.2552	5.0%	0.53 [0.32, 0.87]
Jeremic, 1999	-0.3075	0.1866	7.4%	0.74 [0.51, 1.06]
Luan, 2015	-0.4206	0.1494	9.3%	0.66 [0.49, 0.88]
Luo, 2017	-0.6573	0.2151	6.3%	0.52 [0.34, 0.79]
Qi, 2019	-1.0217	0.2069	6.6%	0.36 [0.24, 0.54]
Qin, 2016	-0.345	0.2429	5.3%	0.71 [0.44, 1.14]
Slotman, 2015	-0.3168	0.0906	13.1%	0.73 [0.61, 0.87]
Xu, 2017	-0.5954	0.1637	8.5%	0.55 [0.40, 0.76]
Xu, 2017	-0.8749	0.3261	3.4%	0.42 [0.22, 0.79]
Zhang, 2019	-0.1128	0.1244	10.8%	0.89 [0.70, 1.14]
Zhu, 2011	-0.5046	0.1499	9.3%	0.60 [0.45, 0.81]
Total (95% CI)			**100.0%**	**0.64 [0.56, 0.72]**

Heterogeneity: Tau² = 0.03; Chi² = 23.36, df = 12 (P = 0.02); I² = 49%
Test for overall effect: Z = 6.72 (P < 0.00001)

C

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
Jeremic, 1999	-0.4734	0.226	23.2%	0.62 [0.40, 0.97]
Luo, 2017	-1.1473	0.2895	18.8%	0.32 [0.18, 0.56]
Qi, 2019	-0.9526	0.2421	22.0%	0.39 [0.24, 0.62]
Xu, 2017	-1.4001	0.2207	23.6%	0.25 [0.16, 0.38]
Xu, 2017	-0.9046	0.4134	12.4%	0.40 [0.18, 0.91]
Total (95% CI)			**100.0%**	**0.38 [0.26, 0.53]**

Heterogeneity: Tau² = 0.09; Chi² = 9.03, df = 4 (P = 0.06); I² = 56%
Test for overall effect: Z = 5.47 (P < 0.00001)

Figure 2 Forest plots of HRs for OS, PFS and LRFS in patients with ES-SCLC. (A) OS; (B) PFS; (C) LRFS. OS, overall survival; PFS, progression-free survival; LRFS, local recurrence-free survival; ES-SCLC, extensive stage small cell lung cancer.
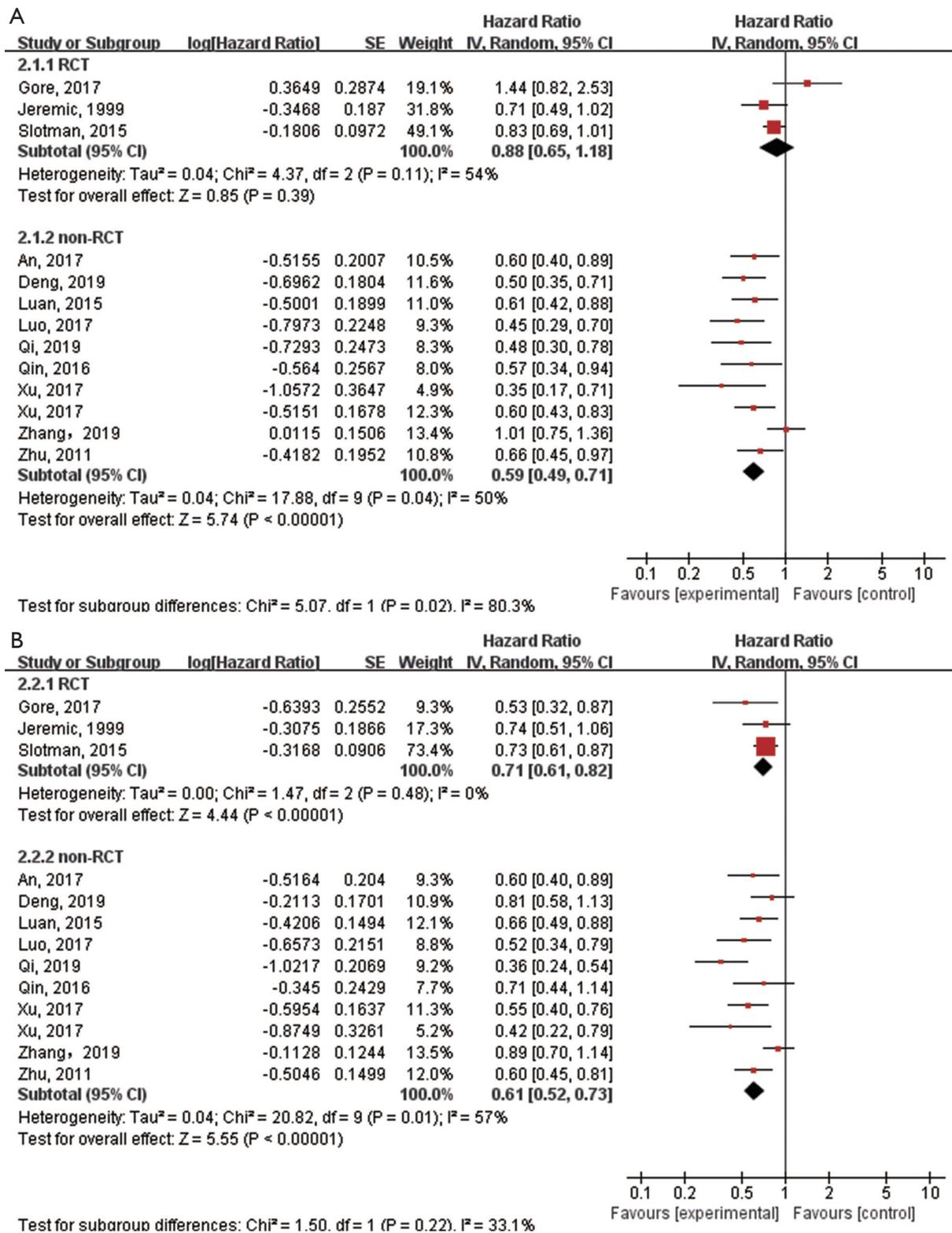
A

B

Figure 3 Forest plots for subgroup analysis of HRs for OS and PFS by study type. (A) OS; (B) PFS. OS, overall survival; PFS, progression-free survival.
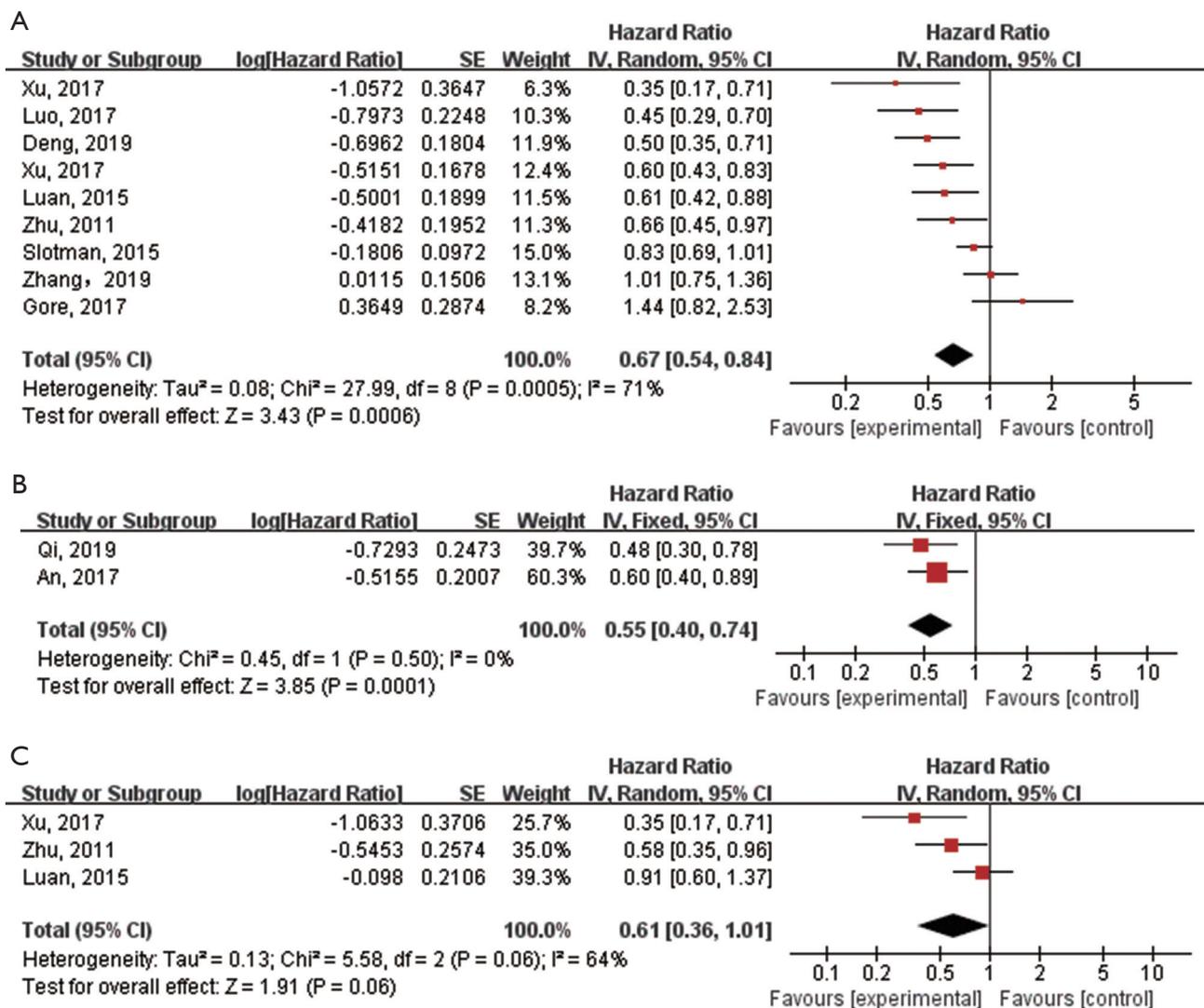
Figure 4 Forest plots of HRs for OS in patients with different clinicopathological features. (A) Patients receiving sequential TRT; (B) patients over 65 years old; (C) patients with oligometastatic disease. OS, overall survival.

Discussion

This meta-analysis compared the clinical efficacy of TRT combined with chemotherapy with chemotherapy alone in ES-SCLC based on 12 studies. The combined results showed that TRT significantly improved both OS and PFS. In addition, TRT reduced the risk of intrathoracic progression and improved LRFS.

Subgroup analysis according to the study type was conducted due to significant heterogeneity of the combined results. The benefits of TRT on OS and PFS were observed in the subgroup of retrospective studies. In the subgroup of RCTs, TRT still improved PFS in patients with ES-SCLC.

However, the pooled analysis of RCTs showed that there was no significant difference in OS between the two groups, which was consistent with the previous meta-analysis (including the same three RCTs) (22). Moderate degree of heterogeneity was observed in the RCT subgroup, which was mainly affected by the negative results of the RTOG-0937 (10). Baseline characteristics of the TRT group and the non-TRT group in RTOG-0937 were unbalanced, and there were more elderly patients (65 years or older) in the TRT group, which may increase the risk of bias. In addition, the TRT group included more patients with poor performance status, 2–4 metastatic lesions, and PR to initial

Page 10 of 13

Li et al. Role of TRT in ES-SCLC
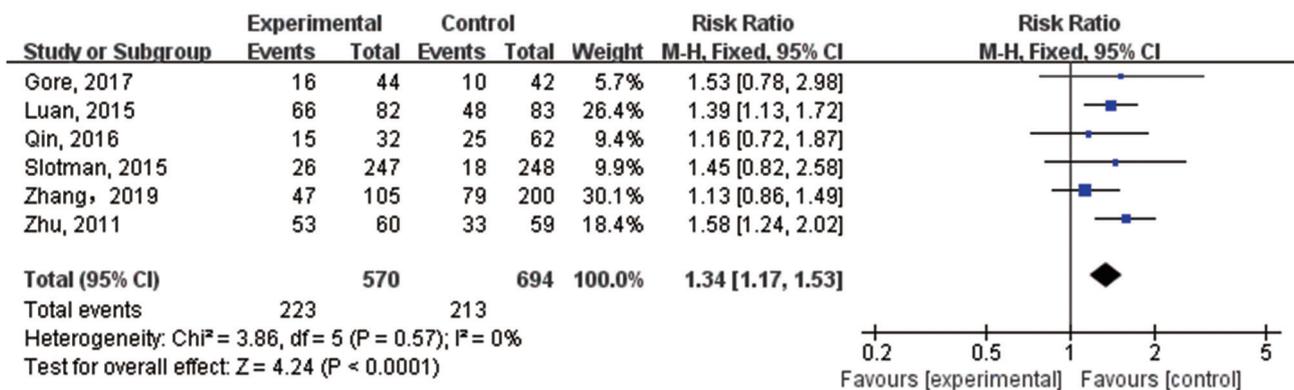
Figure 5 Forest plot of RR for grade III or higher toxicity in patients with ES-SCLC. ES-SCLC, extensive stage small cell lung cancer.

Table 3 Comparison of common grade III or higher toxicity between the TRT group and the non-TRT group

Toxicity (grade ≥3)	No. of studies	TRT group			Non-TRT group			RR (95% CI)	Heterogeneity (I²)	P value
		Events (n)	Total (n)	%	Events (n)	Total (n)	%			
Leucopenia	5	116	273	42.5	98	300	32.7	1.26 (0.82–1.95)	70%	0.29
Thrombocytopenia	5	27	273	9.9	42	300	14	0.67 (0.44–1.03)	0%	0.07
Anemia	5	22	273	8.1	25	300	8.3	0.92 (0.53–1.59)	0%	0.76
Nausea/vomiting	5	16	476	3.4	28	506	5.5	0.61 (0.35–1.08)	46%	0.09
Esophageal toxicity	6	27	581	4.6	0	716	0	13.89 (3.63–53.19)	0%	0.0001
Bronchopulmonary toxicity	7	18	625	2.9	6	748	0.8	2.63 (1.21–5.72)	3%	0.02

Grade ≥3 esophageal toxicity includes: esophagitis, dysphagia; Grade ≥3 bronchopulmonary toxicity includes: dyspnea, cough, pneumonitis. TRT, thoracic radiotherapy; RR, risk ratio.

chemotherapy, which may further weaken the survival advantage of consolidation radiotherapy.

Considering that the survival benefit of platinum-based chemotherapy and concurrent TRT was superior to sequential TRT in patients with limited-stage SCLC, we explored the role of sequential radiotherapy in ES-SCLC (23). Our analysis suggested that sequential TRT had better OS than the control group. The effect of concurrent TRT in ES-SCLC was not evaluated due to insufficient data. Of the twelve included studies, only one randomized study compared survival in patients receiving concurrent TRT with those receiving chemotherapy alone. The results showed that concurrent TRT provided a significant survival benefit (5).

In the meta-analysis, we further explored the role of TRT in patients with ES-SCLC with different clinicopathologic features. The pooled results showed that the addition of TRT still improved the prognosis of patients over 65 years

of age. In addition, previous studies have demonstrated that ES-SCLC patients with a single metastatic site were more suitable for radiotherapy as they had lower disease burden (16,24,25). However, our study did not observe the benefit of TRT for OS in patients with only one organ metastasis. Since only three studies have reported OS in patients with oligometastatic disease and they were all retrospective studies, this conclusion should be considered with caution. The study of Zhang *et al.* was not included in this subgroup analysis because they excluded patients with brain or liver metastases from the oligometastatic population. In their study, TRT significantly increased PFS and OS in patients with oligometastatic ES-SCLC without brain or liver metastasis (11). It is worth noting that a secondary analysis of the CREST trial found that both OS and PFS benefits were observed in patients with two or fewer metastases (no brain metastasis). No difference in survival was observed between patients with 0–1 metastatic sites and 2 metastatic
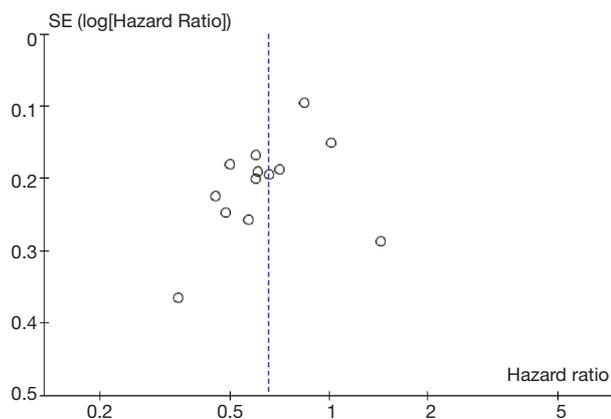
Figure 6 Funnel plot for publication bias.

sites (25). Moreover, in the first randomized controlled study that found the survival benefits of TRT in ES-SCLC, most patients (about 90%) had two or fewer metastases (no brain metastasis) (5). Therefore, the definition of oligometastases (with or without brain metastasis, one or two metastasis sites) in future studies still needs to be unified. Because of the limited number of studies, we did not conduct a subgroup analysis of survival in patients with multiple metastases. A recent retrospective study found that TRT also significantly improved prognosis in patients with multiple metastases (18). Another study showed that there was no statistical difference in PFS and OS between TRT and non-TRT groups for patients with brain/liver/multi-metastasis (11). In the future, further prospective RCTs are needed to assess the role of TRT in ES-SCLC patients with different numbers of metastatic sites. As far as we know, our study is the first to report these results.

Our meta-analysis suggested that patients receiving consolidation TRT had a higher risk of grade ≥3 toxicity. We then compared the common grade ≥3 toxicity between the TRT group and the control group, and found no statistical difference in hematologic toxicity (leukopenia, thrombocytopenia, anemia) and non-hematologic toxicity (nausea or vomiting) between the two groups. Consolidation TRT resulted in a higher incidence of grade III or higher esophageal and bronchopulmonary toxicity than the non-TRT group. However, the incidence of grade ≥3 esophageal (4.6%) and bronchopulmonary (2.9%) toxicity in TRT group was low. Furthermore, radiation-induced deaths were infrequent, with only two treatment-related deaths in all included studies. They all died of radiation pneumonitis (10,21). In conclusion, chemotherapy combined with TRT was well tolerated in patients with ES-SCLC.

There are some limitations in our study. First, the survival data in some studies were estimated from Kaplan-Meier curves, which may reduce the accuracy of the results. Second, our study included three RCTs and nine retrospective studies. Retrospective studies may introduce selection bias. Third, moderate degree of heterogeneity was found in some combined analyses (all $I^2<75\%$). This may be due to the differences in baseline characteristics (performance score, brain metastasis, disease burden), chemotherapy-related issues (response to initial chemotherapy, chemotherapy cycle), radiotherapy regimen (radiation dose, radiotherapy timing), and PCI or not in the included study. Subgroup analyses were not conducted due to insufficient information on these factors in the included studies. However, we performed the sensitivity analysis and found that the benefits of TRT on OS, PFS and LRFS were not affected after excluding any single study, which confirmed the stability of our results.

Conclusions

Our meta-analysis showed that TRT improves OS, PFS and LRFS in patients with ES-SCLC. Further RCTs are expected to confirm our conclusions.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at http://dx.doi.

Page 12 of 13

Li et al. Role of TRT in ES-SCLC

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-20-5765). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Rudin CM, Giaccone G, Ismaila N. Treatment of Small-Cell Lung Cancer: American Society of Clinical Oncology Endorsement of the American College of Chest Physicians Guideline. J Oncol Pract 2016;12:83-6.
2. Parikh M, Riess J, Lara PN Jr. New and emerging developments in extensive-stage small cell lung cancer therapeutics. Curr Opin Oncol 2016;28:97-103.
3. Slotman B, Faivre-Finn C, Kramer G, et al. Prophylactic cranial irradiation in extensive small-cell lung cancer. N Engl J Med 2007;357:664-72.
4. Alvarado-Luna G, Morales-Espinosa D. Treatment for small cell lung cancer, where are we now?-a review. Transl Lung Cancer Res 2016;5:26-38.
5. Jeremic B, Shibamoto Y, Nikolic N, et al. Role of radiation therapy in the combined-modality treatment of patients with extensive disease small-cell lung cancer: A randomized study. J Clin Oncol 1999;17:2092-9.
6. Slotman BJ, van Tinteren H, Praag JO, et al. Use of thoracic radiotherapy for extensive stage small-cell lung cancer: a phase 3 randomised controlled trial. Lancet 2015;385:36-42.
7. Zhang R, Li P, Li Q, et al. Radiotherapy improves the survival of patients with extensive-disease small-cell lung cancer: a propensity score matched analysis of Surveillance, Epidemiology, and End Results database. Cancer Manag Res 2018;10:6525-35.
8. Deng L, Zhou Z, Xiao Z, et al. Impact of thoracic radiation therapy after chemotherapy on survival in extensive-stage small cell lung cancer: A propensity score-matched analysis. Thorac Cancer 2019;10:799-806.
9. Qin T, Zhou N, Zeng YD, et al. Benefit from thoracic radiotherapy in patients with extensive-disease small-cell lung cancer with elevated lactate dehydrogenase. Onco Targets Ther 2016;9:1095-103.
10. Gore EM, Hu C, Sun AY, et al. Randomized Phase II Study Comparing Prophylactic Cranial Irradiation Alone to Prophylactic Cranial Irradiation and Consolidative Extracranial Irradiation for Extensive-Disease Small Cell Lung Cancer (ED SCLC): NRG Oncology RTOG 0937. J Thorac Oncol 2017;12:1561-70.
11. Zhang H, Deng L, Wang X, et al. Metastatic location of extensive stage small-cell lung cancer: implications for thoracic radiation. J Cancer Res Clin Oncol 2019;145:2605-12.
12. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol 2010;25:603-5.
13. Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.
14. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17:2815-34.
15. Tierney JF, Stewart LA, Ghersi D, et al. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007;8:16.
16. Qi J, Xu L, Sun J, et al. Thoracic Radiotherapy Benefits Elderly Extensive-Stage Small Cell Lung Cancer Patients with Distant Metastasis. Cancer Manag Res 2019;11:10767-75.
17. Luo J, Xu L, Zhao L, et al. Timing of thoracic radiotherapy in the treatment of extensive-stage small-cell lung cancer: important or not? Radiat Oncol 2017;12:42.
18. Xu LM, Cheng C, Kang M, et al. Thoracic radiotherapy (TRT) improved survival in both oligo- and polymetastatic extensive stage small cell lung cancer. Sci Rep 2017;7:9255.
19. An C, Jing W, Zhang Y, et al. Thoracic radiation therapy could give survival benefit to elderly patients with extensive-stage small-cell lung cancer. Future Oncol 2017;13:1149-58.
20. Luan Z, Wang Z, Huang W, et al. Efficacy of 3D

conformal thoracic radiotherapy for extensive-stage small-cell lung cancer: A retrospective study. Exp Ther Med 2015;10:671-8.

21. Zhu H, Zhou Z, Wang Y, et al. Thoracic radiation therapy improves the overall survival of patients with extensive-stage small cell lung cancer with distant metastasis. Cancer 2011;117:5423-31.

22. Rathod S, Jeremic B, Dubey A, et al. Role of thoracic consolidation radiation in extensive stage small cell lung cancer: A systematic review and meta-analysis of randomised controlled trials. Eur J Cancer 2019;110:110-9.

23. Takada M, Fukuoka M, Kawahara M, et al. Phase III study of concurrent versus sequential thoracic radiotherapy in combination with cisplatin and etoposide for limited-stage small-cell lung cancer: results of the Japan Clinical Oncology Group Study 9104. J Clin Oncol 2002;20:3054-60.

24. Fukui T, Itabashi M, Ishihara M, et al. Prognostic factors affecting the risk of thoracic progression in extensive-stage small cell lung cancer. BMC Cancer 2016;16:197.

25. Slotman BJ, Faivre-Finn C, van Tinteren H, et al. Which patients with ES-SCLC are most likely to benefit from more aggressive radiotherapy: A secondary analysis of the Phase III CREST trial. Lung Cancer 2017;108:150-3.