

- pelvic lymph node metastasis in early-stage cervical cancer. *Eur Radiol.* 2021;31:5967–5979.
- Hu Y, Xie C, Yang H, et al. Assessment of intratumoral and peritumoral computed tomography radiomics for predicting pathological complete response to neoadjuvant chemoradiation in patients with esophageal squamous cell carcinoma. *JAMA Netw Open.* 2020;3:e2015927.
 - Xu X, Zhang HL, Liu QP, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol.* 2019;70:1133–1144.
 - Bera K, Braman N, Gupta A, et al. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol.* October 18, 2021 [Epub ahead of print].
 - Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14:749–762.
 - Liang W, Yang P, Huang R, et al. A combined nomogram model to preoperatively predict histologic grade in pancreatic neuroendocrine tumors. *Clin Cancer Res.* 2019;25:584–594.

Lu Zhang
Bin Zhang*

*First Affiliated Hospital of Jinan University
Guangdong, Guangzhou, China
E-mail: xld_Jane_Eyre@126.com

Published online Dec. 16, 2021.
DOI: 10.2967/jnumed.121.263479

Reply: Adding Nontumor Radiomic Features to the Prognostic Model Is Bothersome but Useful

REPLY: We appreciate the comments (1) by Zhang and Zhang in response to our manuscript (2). We thank them for their compliments about our work overall. However, we would also like to clarify several points of confusion for their sake and for readers generally.

First, our feature selection involved more than simply “manually excluding several highly correlated features.” We also performed feature reduction by individually assessing features for significant association with recurrence, using bootstrap resampling. We then performed forward stepwise feature selection for model building, carefully checking for nonredundancy of added features at each step, using model stability testing based on bootstrap analysis of the out-of-bag C-indices. Consequently, our stepwise selections terminated after no more than 5 rounds, instead of spuriously adding correlated features. Additionally, to further test that these selected features were not overfitting to noise, we evaluated our model on the test set and calculated C-index CIs to ensure that training and test C-indices overlapped. This combination of procedures more than satisfies the authors’ demand for “more sophisticated and rigorous dimensionality reduction methods . . . to ensure the reproducibility and independence

of the identified radiomic features” (1). We also note that the rigorous tests for nonredundancy that we implemented rebut the authors’ comments regarding subgroup analysis by stage, because the nonredundancy between radiomics and stage is already built-in to the model development.

Second, regarding the statement that “Figure 6 showed the same hazard ratios in the models . . . suggesting that whole-body biomarkers failed to provide additional information for risk stratification.” We presume the authors are referring to the comparison between Figures 6B and 6C. However, as we mention, the risk score cutoff in Figure 6C was explicitly chosen solely to ensure an equal number of high-risk and low-risk cases as in Figures 6A and 6B for comparison. It was not chosen to optimize the hazard ratio; however, the cutoff in Figure 6D was chosen in this way, where one can see that the optimal stratification using whole-body radiomics-based risk score outperforms the other stratifications.

We agree with some of the authors’ points. Although we discussed potential physiologic correlates of our features, the biologic meaning of the features would undoubtedly be more precise by “correlation with computational pathology features, radiology–pathology coregistration, or analysis of biologic pathways or genomic correlations.” This was beyond our scope of work but is a worthy line of inquiry for future validation studies. We also agree that other measures of model performance could have been reported, including decision curve analysis, calibration plots, or net reclassification improvement. Indeed, there are a wide variety of measures that are reasonable, and the choice of which to report always involves an element of arbitrariness. The metrics we reported, including the C-indices, 2-y receiver-operating-characteristic curve and risk stratifications with hazard ratio, were chosen on the basis of their widespread usage in the biostatistics literature. Nevertheless, we concede that future validation studies would benefit from a more comprehensive set of assessments.

REFERENCES

- Zhang L, Zhang B. Adding nontumor radiomic features to the prognostic model is bothersome but useful [letter]. *J Nucl Med.* 2022;63:494–495.
- Yusufaly T, Zou J, Nelson T, et al. Improved prognosis of treatment failure in cervical cancer with non-tumor positron emission tomography/computed tomography radiomics. *J Nucl Med.* October 28, 2021 [Epub ahead of print].

Tahir I. Yusufaly*
Loren K. Mell

*Johns Hopkins University, School of Medicine
Baltimore, Maryland
E-mail: tyusufa2@jhmi.edu

Published online Jan. 13, 2022.
DOI: 10.2967/jnumed.121.263730