We acknowledge the concern of Banerjee et al. (*1*), as it becomes clear from our respective remarks concerning this limitation in the introduction and discussion sections. In fact, we have concluded that exendin PET has the potential to be an improvement over ^{18}F-DOPA PET, but that more work is needed to validate these preliminary results (*2*).

To allay the concerns of bias spilling into the results of our study, we would like to stress that our group did realize that PET readings could be positively biased if the reporting was left only to the inventors of the technique. That is why we chose 2 methods of reading—that is, clinical reading at the respective site where imaging had been performed and an independent, masked reading. Thereafter, to overcome any discordant results, a joint reading was performed. Although we took the joint reading as reference, knowing that the gold standard for imaging findings can only be histopathology, specifically for new tracers and new indications, we did perform histopathologic confirmation in all specimens operated on (Supplemental Table 2 in (*2*)). Even the surgeons, trained on using ^{18}F-DOPA PET–directed surgical planning, clearly stated more confidence in exendin PET in interpretation of the image results than in ^{18}F-DOPA PET. These results are presented in the supplemental materials using the validated Likert scale (*2*).

We are grateful for the comments from Banerjee et al. supporting the conclusions we have drawn, as the points they raised are in line with the arguments we have put forward in our paper. Also, we are grateful that Banerjee et al. raise the issues for which the answers can be found in the supplemental material, thus putting emphasis on this part of the paper as well. In addition, we hope that our remarks with respect to a potential bias have helped to allay such concerns.

REFERENCES

1. Banerjee I, Sajjan R, Salomon Estebanez M, et al. ^{68}Ga-NODAGA-exendin-4 PET scanning for focal congenital hyperinsulinism: need for replication [letter]. *J Nucl Med.* 2021;63:493.
2. Boss M, Rottenburger C, Brenner W, et al. ^{68}Ga-NODAGA-exendin-4 PET improves the detection of focal congenital hyperinsulinism. *J Nucl Med.* July 2, 2021 [Epub ahead of print].

Vikas Prasad
Marti Boss
Christof Rottenburger
Winfried Brenner
Oliver Blankenstein
Sonal Prasad
Mijke Buitinga
Pratik Shah
Martin Gotthardt*
**Radboud University Medical Centre*
Nijmegen, The Netherlands
E-mail: martin.gotthardt@radboudumc.nl

DOI: 10.2967/jnumed.121.263167

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

TO THE EDITOR: We read with great interest the article by Dr. Yusufaly and colleagues (*1*), who developed a radiomic model incorporating tumor radiomic features, nontumor radiomic features, and clinical variables to predict disease recurrence in patients with cervical cancer. To the best of our knowledge, this was the first study to suggest that nontumor radiomic biomarkers derived from the whole body (including bone, fat, and muscle) could improve prognostic modeling of cancer. Previous studies have proven that peritumoral radiomic features could improve the performance of radiomic models (*2–4*). This study provides more comprehensive insights into the tumor and the immune state of the human body.

Despite the encouraging results, several methodologic issues should be noted. First, we are concerned about the workflow of the radiomic analysis. Although this study evaluated its radiomic quality score of 18 points (total points of 36), 2 domains were not truly conducted, that is, detection and discussion of biologic correlates and potential clinical utility. Although this study hypothesized that whole-body radiomic features may be associated with immune system function and could reflect variation in patients' global inflammatory state, it did not investigate the biologic meaning behind radiomic features by correlation with computational pathology features, radiology–pathology coregistration, or analysis of biologic pathways or genomic correlations (*5*). In addition, an assessment of potential clinical utility through statistical methods such as decision curve analysis was not performed. Second, the current feature selection is not enough despite the fact that the authors manually excluded several highly correlated features; more sophisticated and rigorous dimensionality reduction methods (such as intraclass correlation analysis and Pearson correlation coefficient analysis) must be implemented to ensure the reproducibility and independence of the identified radiomic features (*6*). Third, this study applied only C-index as a discrimination metric for evaluating the predictive performance of radiomic models, but this metric is not enough, as calibration is not fully captured by C-index. Calibration statistics such as calibration plots, which reflect the consistency between the true probability and the predicted probability, are needed (*7*). Both discrimination and calibration statistics are recommended when evaluating the performance of models. Fourth, use of the cindex.comp package and net reclassification improvement is recommended for pairwise comparisons of model performance. Fifth, given the distinct prognosis between early-stage and advanced-stage tumors, the risk stratification determined by radiomics may be confounded by tumor stage; a subgroup analysis by stage can be considered to identify the true effect of radiomics. In addition, Figure 6 showed the same hazard ratios in the models based on stage plus tumor-related biomarkers and the model based on all biomarkers, suggesting that whole-body biomarkers failed to provide additional information for risk stratification. Finally, as the authors acknowledged in the limitations, the radiomic model was developed and validated at a single small center; multiple external validations would be beneficial for more generalizability to heterogeneous groups of patients regardless of the clinical setting.

Despite the above-mentioned limitations, we still appreciate Yusufaly and colleagues for their outstanding work on nontumor radiomic biomarker analysis, which provides a more holistic model. We look forward to further works to improve the validity and generalizability of their radiomic models.

REFERENCES

1. 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].
2. Li XR, Jin JJ, Yu Y, Wang XH, Guo Y, Sun HZ. PET-CT radiomics by integrating primary tumor and peritumoral areas predicts E-cadherin expression and correlates with

pelvic lymph node metastasis in early-stage cervical cancer. *Eur Radiol.* 2021;31:5967–5979.
3. 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.
4. 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.
5. 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].
6. 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.
7. 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

1. Zhang L, Zhang B. Adding nontumor radiomic features to the prognostic model is bothersome but useful [letter]. *J Nucl. Med.* 2022;63:494–495.
2. 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