AWARD NUMBER: W81XWH-15-1-0110

TITLE: Noninvasive Characterization of Indeterminate Pulmonary Nodules Detected on Chest High-Resolution Computed Tomography

PRINCIPAL INVESTIGATOR: Fabien Maldonado, M.D.

CONTRACTING ORGANIZATION: Vanderbilt University Medical Center, Nashville, TN

REPORT DATE: October 2021

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command

Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| **1. REPORT DATE** | **2. REPORT TYPE** | **3. DATES COVERED** |
|---|---|---|
| October 2021 | Annual | 30Sep2020-29Sep2021 |

**4. TITLE AND SUBTITLE**

Noninvasive Characterization of Indeterminate Pulmonary Nodules Detected on Chest High-Resolution Computed Tomography

**5a. CONTRACT NUMBER**
W81XWH-15-1-0110

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Fabien Maldonado, M.D.

E-Mail:

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

VANDERBILT UNIVERSITY MEDICAL CENTER
VUMC
1161 21ST AVE S STE D3300 MCN
NASHVILLE TN 37232-0011

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Lung cancer accounts for more cancer-related mortality than breast, prostate and colon cancer combined, and greater than half of lung cancer cases are diagnosed at advanced stages. Both the National Lung Screening Trial (NLST) and Nederlands-Leuvens Longkanker Screenings ONderzoek (NELSON) trial have demonstrated that low dose CT screening (LDCT) reduces mortality in high-risk patients. Implementation of LDCT screening has however been challenging.

**15. SUBJECT TERMS**
None listed.

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | USAMRMC |
| Unclassified | Unclassified | Unclassified | Unclassified | 20 | **19b. TELEPHONE NUMBER** *(include area code)* |

## Table of Contents


1. **Introduction – 3**
2. **Keywords – 4**
3. **Accomplishments – 4**
4. **Impact – 15**
5. **Changes/Problems – 16**
6. **Products – 17**
7. **Participants & Other Collaborating Organizations – 18**
8. **Special Reporting Requirements – 18**
9. **Appendices – 19**

## 1. INTRODUCTION:

Lung cancer accounts for more cancer-related mortality than breast, prostate and colon cancer combined, and greater than half of lung cancer cases are diagnosed at advanced stages. Both the National Lung Screening Trial (NLST) and Nederlands-Leuvens Longkanker Screenings ONderzoek (NELSON) trial have demonstrated that low dose CT screening (LDCT) reduces mortality in high-risk patients. Implementation of LDCT screening has however been challenging. A main reason for the low uptake of LDCT screening is the high false positive rate, and the uncertainty about optimal interval between screening rounds and cost-effectiveness. The high false positive rates noted in NLST has been improved by the application of the Lung Imaging Reporting and Data System (Lung-RADS) criteria, where smaller and part-solid nodules (<6mm), nodules with specific calcification patterns, perifissural nodules (<10 mm) are classified as benign. While Lung-RADS reduces the false positive rate to 5.3%, it also reduces sensitivity at baseline by about 10%. To put things in context, even applying the Lung-RADS criteria to 1000 patients in NLST, in an effort to prevent 3 cancer related deaths, 779 would have had normal scans, 180 who do not have cancer would have gotten an extra scan and 13 of those would have needed an unnecessary invasive procedure. This problem is compounded by the fact that incidentally identified lung nodules are also on the rise, presumably due to increased use of high-resolution CT (HRCT). Based on data from 2006 to 2012, it is estimated that around 1.5 million adult Americans will have a pulmonary nodule identified each year, which is almost certainly an underestimate. Because the vast majority of these incidentally- or screen-identified lung nodules will ultimately prove benign, efficient diagnostic investigations will be increasingly important and while many biomarkers are currently being evaluated to optimize nodule management, none has been widely adopted in clinical practice. Risk-stratification of nodules as allowed by current clinical prediction models remain suboptimal, and the development and validation of cost-effective tools to guide management of larger (≥ 7 mm) screen-detected nodules is needed to mitigate the problem of unnecessary diagnostic interventions leading to excessive mortality, morbidity and healthcare costs.

Quantitative radiologic analysis of available CT datasets using artificial intelligence (radiomics) is an attractive option, leveraging existing datasets, obviating the need for additional invasive or non-invasive investigations. Most recent approaches have focused on deep learning methods, which have shown promises in many other fields, with the caveat that predictive variables thus identified are without explicit correlation to tumor biology, and that deep learning methods require large curated and annotated datasets for development and validation which are not easily accessible. In addition, there is great variability in image acquisition, feature extraction and statistical analysis of the various radiomic models described in literature. While many promising models have been developed, external validations are rare, a consequence of the paucity of available curated datasets and the risk of overfitting that continues to plague radiomic models, particularly those relying on deep-learning methods. It is also often unclear whether such models outperform validated simpler and readily accessible clinical prediction models.

We have previously described the development of a radiomic classifier that effectively distinguishes benign from malignant nodules using a training set of 726 indeterminate nodules from the NLST database, which was validated with relatively good diagnostic test performance on two independent datasets with large prevalence of malignancy (the early available DECAMP-1 dataset (which has since stopped accruing new patients) and the Lung Tissue Registry Consortium dataset with acceptable performance (0.66 and 0.80 AUC, respectively)). In the past year, we validated our model using a curated independent dataset from the Vanderbilt University lung nodule registry, and used clinical and radiologic variables to calculate the probability of malignancy of the NLST and Vanderbilt University registry datasets based on a commonly used

and validated clinical prediction rule, the Brock model, in order to estimate the probability of reclassification of screen- or incidentally-identified lung nodules with intermediate probability of malignancy (defined as 10% to 60% as calculated by a validated predictive model, the Brock model) into classes of low- or high-probability of malignancy. This work has since the last annual report been published in the European Respiratory Journal. We have also since resubmitted an application to the DECAMP investigators to use the now completed DECAMP-1 dataset for validation of our model and have received the images (as of March 2021). Validation using the DECAMP-1 dataset will be the object of this coming year's proposed work as part of a no-cost extension. While this work was proposed to be completed last year, two unexpected major events prevented us from completing this work and motivated a request for another no-cost extension: (1) the COVID-19 pandemic, which placed a disproportionate burden on our division of pulmonary and critical medicine and slowed down our work considerably, and (2) the sudden and unexpected passing of Pierre Massion, MD, on April 4th, 2021, who led the Massion Biomarker Laboratory at Vanderbilt University Medical Center, in which image analysis was to be performed, required us to modify our plans and transfer the images to the Biomedical Imaging Resource laboratory at Mayo Clinic, Rochester, MN, for further image segmentation and analysis. These datasets have now been successfully transferred.

## 2. KEYWORDS:

Lung cancer, Radiomics, Lung cancer screening, Chest computed tomography, Biomarker, Lung nodules.

## 3. ACCOMPLISHMENTS:

**What were the major goals of the project?**

**Aim 1 (first year of the grant):** The first aim of this grant was to develop a CT-based radiomic model using quantitative volumetric analysis of screen-identified lung nodules (model 1), and a combined clinical-radiologic model (model 2) to differentiate benign from malignant nodules.

a. Milestone: Development of optimized quantitative radiological variables predictive of the benign or malignant character of lung nodules from a cohort isolated from the NLST (12 months – October 2016).

   *Note that subcontracts with Brown University and Mayo Clinic (required due to relocation of the PI, Fabien Maldonado, to Vanderbilt University) were not established until March 2016 and as such work could not be started before that time.*

   The identification of optimization of quantitative radiological variables was completed by October 2017.

b. Milestone: development of a radiologic prediction model (12 months)

   The radiologic model was completed by October 2017.

c. Development of a combined clinical/radiologic prediction model (12 months).

The clinical/radiologic model was completed by October 2017, but addition of clinical variables did not contribute significantly to the diagnostic test performance of the model and therefore was not selected as the final predictive model in subsequent validations.

**Aim 2 (second year of the grant):** the second aim of this grant is to prospectively validate the models developed in Aim 1 in the DECAMP-1 dataset (500 patients with indeterminate pulmonary nodules**,** DECAMP PROTOCOL ACRIN 4703).

1. Milestone: Validation of a radiologic and combined clinical/radiologic prediction models (**Year 2 of the grant**).

   Enrollment for the DECAMP-1 study was considerably delayed and eventually halted before the expected enrollment of 500 patients at the end of February 2020. This completion of enrollment was anticipated by December 2015 at the time of our application (August 2014), as 125 of the planned 500 patients had already been enrolled (see attached original support letter from DECAMP-1 PI Dr. Avrum Spira, Boston University). As of October, 2021: 489 patients have been recruited to DECAMP-1 (enrollment completed) and nodules have been adjudicated (personal communication, Fenghai Duan, co-I). As previously reported, an interim blind analysis of this dataset that included 274 nodules (183 malignant and 91 benign, confirmed) and using our radiologic model yielded an AUC of 0.66 (strict validation) and 0.74 (loose validation). These results were felt to be due to the very large number of malignant nodules in this cohort, likely to result in early adjudication of the most suspicious lung nodules. Contractual agreement between the American college of Radiology and Vanderbilt University Medical Center was established and we expect finalization of the analysis in the next quarter.

   We were however able to secure two alternative validation sets, including (1) a validation cohort from the lung nodule registry at Vanderbilt University Medical Center/Nashville Veterans Administration Tennessee Valley Healthcare system (primary investigator: Dr. Pierre Massion, see below) comprised of 91 malignant lung nodules and 79 benign lung nodules (total n=170), as well as (2) the lung nodule cohort from the Lung Tissue Research Consortium (LTRC), comprised of 88 benign and 89 malignant nodules (total n=177). Similar to the early DECAMP-1 cohort, the LTRC cohort was considered “high-risk” as all nodules were evaluated by expert radiologists and felt to be suspicious enough for malignancy to require surgical resection (see below). The validation from the LTRC dataset was reported in a previous report and is summarized below. We are in the process of curating another validation dataset of high-risk nodules from the Vanderbilt University TREAT dataset, which contains 600 surgically resected nodules (200 benign and 400 malignant) to further validate our model on a challenging dataset. Note that funds from the award have not and are not being used for these external validation sets. The no-cost extension was solely requested to complete the proposed project to validate our mode on the DECAMP-1 dataset.

**What was accomplished under these goals?**

a. **Major activities**
   Summary of activities that occurred during the first year of the grant:

Year 1 (for additional details see prior annual report):

Nodule selection

As previously reported, participants for our project were selected from the pool of eligible participants in the NLST, who did not withdraw from follow-up, in the CT arm of the study (N=26,262) and included all screen-detected lung cancer cases. Non-lung cancer controls were selected as a stratified random sample from all participants in the pool defined above who were not found to have lung cancer during the screen or follow-up periods of the NLST in a 1:1 fashion. We restricted our analysis to nodules with a size defined by a largest diameter comprised between 7 and 30 mm as reported in the NLST database, as these represent the size criteria used as eligibility criteria in the DECAMP-1 study

CT dataset image transfer, segmentation and analysis have been previously reported.

Nodule segmentation and analysis

The lung nodules were segmented manually using the ANALYZE software (Biomedical Imaging Resource, Mayo Clinic, Rochester, MN), as previously reported. A comprehensive set of automatically computable, quantitative radiomic metrics was included for the development of a multivariable predictive model to discriminate benign from malignant lung nodules. Based on previous data and preliminary analysis, we considered metrics within the following categories: general characteristics of the nodule (volume and location), nodule characteristics (texture and surface characteristics) and nodule-free surrounding lung characteristics.

Year 2 (for additional details see prior annual report):

Multivariate analysis was performed using least absolute shrinkage and selection operator (LASSO) method for both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the multivariate statistical model. To increase the stability of the modeling, LASSO was run 1,000 times and the variables that were selected by at least 50% of the runs were included into the final multivariate model. The bootstrapping method was then applied for the internal validation, and the optimism-corrected AUC was reported for the final model.

Results:

After exclusion of cases lacking HRCT data, cases with no apparent lesion on last HRCT prior to the cancer diagnosis, cases with nodules invading the mediastinum, cases with missing outcome data, and lesion with size < 7mm or >30 mm, 408 LDCT scans with malignant nodules were selected and analyzed. A stratified random sample of non-lung cancer controls (nodules with size comprised between 7 and 30 mm) was selected on a 1:1 basis, and after exclusion of HRCT containing more than one nodule, 318 nodules were selected and included in the analysis.
The demographic and clinical characteristics of individuals included in the study are summarized as follows:

**Demographics and Clinical Characteristics of Cancer and Control (n = 726)**

| | Lung Cancer Cases (n=408) | Nodule-Positive Controls (n=318) | p Value |
|---|---|---|---|
| **Age, mean ± SD, y** | 63 .7 ± 5.3 | 61.2 ± 5.0 | <0.001 |
| **Sex, n (%)** | | | 0.45 |
| **Male** | 230 (56.4) | 189 (59.4) | |
| **Female** | 178 (43.6) | 129 (40.6) | |
| **Race, n (%)** | | | 0.03 |
| **White** | 385 (94.4) | 286 (89.9) | |
| **Black, Asian, other** | 23 (5.6) | 32 (10.1) | |
| **Ethnicity, n (%)** | | | 0.31 |
| **Hispanic or Latino** | 405 (98.4) | 313 (99.3) | |
| **Neither Hispanic nor Latino** | 3 (1.6) | 5 (0.7) | |
| **Smoking, n (%)** | | | 0.37 |
| **Current** | 221 (54.2) | 161 (50.6) | |
| **Former** | 187 (45.8) | 157 (49.4) | |
| **Pack-years smoked, mean ± SD** | | | |
| **Current smokers** | 64.8 ± 25.8 | 55.5 ± 20.9 | <0.001 |
| **Former smokers** | 66.7 ± 30.6 | 55.2 ± 26.9 | <0.001 |
| **Self-reported history of COPD, n (%)** | | | |
| **Yes** | 43 (10.5) | 18 (5.7) | 0.02 |
| **No** | 365 (89.5) | 300 (94.3) | |
| **FH of lung cancer, n (%)** | | | 0.08* |
| **Yes** | 113 (28.9) | 69 (22.8) | |
| **No** | 278 (71.1) | 233 (77.2) | |
| **Missing** | n=17 | n=16 | |
| **Stage, n (%)** | | | — |
| **I** | 298 (73.0) | — | |
| **II** | 29 (7.1) | — | |
| **III** | 55 (13.5) | — | |
| **IV** | 20 (5.0) | — | |
| **Carcinoid, unknown** | 6 (1.5) | — | |
| **Histologic subtype, n (%)** | | | — |
| **Adenocarcinoma** | 290 (71.1) | — | |
| **Squamous cell carcinoma** | 81 (19.9) | — | |
| **Other, NOS, unknown** | 37 (9.1) | — | |

P Values calculated using Fisher's exact test for categorical variables, Student's t test for continuous variables.
* P value for family history of lung cancer was calculated without missing data.

Multivariate analysis

In order to select the optimal variables across a set of pre-selected 57 variables previously shown to be predictive of malignancy, adjust the regression coefficients to optimize the transportability

(external validity) of the model and determine the degree of optimism of the model and perform optimism-corrected analysis of the performance of the model by ROC analysis, all 57 variables were included in the LASSO regression model. Multivariate analysis using LASSO on all features yielded a multivariate model with 8 selected features (selected with frequency > 50% after introducing bootstrap to reduce variability after 1000 runs) with an AUC estimate of 0.941. These 8 features include: 1. centroid_Z, 2. Min Enclosing Brick, 3. flatness, 4. SILA_Tex, 5. Max_SI, 6. Avg_SI, 7. Avg_PosMeanCurv and 8. Min_MeanCurv, all with P<0.01. To correct overfitting (internal validation) we used the bootstrapping technique to estimate the optimism of the AUC. **The optimism-corrected AUC is 0.939**.

**Centroid_z** captures the location of the nodule in the lung (vertical axis), the **minimal enclosing brick** and **flatness** capture volume and shape, respectively, **Sila_Tex** is a summary variable capturing the degree of abnormality based on texture density within the nodule, **maximum** and **average shape index (Max_SI and Avg_SI)** capture the complexity of the nodule surface and **Average positive mean curvature and (Avg_PosMeanCurv)** and **Minimum mean curvature (Min_MeanCurv)** represents the degree of curvature of the outer surface of the nodule.

We then added variables independently associated with an increased risk of lung cancer in our cohort (age and pack-years). The optimism-corrected AUC for model 2 is 0.941, not significantly different that model 1 (radiomic model).

***Validation:***

1. DECAMP-1 (see previous report for details):

Due to considerable delay in enrollment of the DECAMP1 study (see above), validation of our model on a prospective cohort of screened individuals similar to those enrolled in the NLST is still pending. Access to this dataset was granted by the DECAMP biomarker committee and image transfer completed in March 2021. As previously reported, an interim analysis that included 274 nodules (183 malignant and 91 benign, confirmed), yielded an AUC of 0.66 (strict validation) and 0.74 (loose validation).

a. Strict DECAMP-1 validation (i.e., using the 8-feature logistic model developed from the NLST data to impute the probability of lung cancer occurrence for each DECAMP case): AUC: 0.6567

b. Loose DECAMP-1 validation (i.e., using the same 8 features identified from the NLST data to then re-fit the logistic regression, plus Bootstrap correction): AUC: 0.7415

As we are proceeding with analysis of the now available DECAMP1 dataset, alternative validation datasets were pursued, curated and analyzed. *Note that no funding from this grant was used for these analyses*.

2. Lung Tissue Research Consortium validation (see previous report for details):

The radiomic model was validated using the Lung Tissue Research Consortium dataset, comprised of 88 benign and 89 malignant nodules. This cohort was considered “high-risk” as all nodules in this cohort were evaluated by expert radiologists and felt to be suspicious enough for

malignancy to require surgical resection (i.e. a nodules, benign and malignant, were resected lung nodules and therefore with a high pre-test probability than typical screen- or incidentally identified lung nodules).

Using these 177 nodules, the results were as follow:

**Sensitivity: 87.6%**

**Specificity: 68.2%**

PPV: 73.6%

NPV: 84.5%

Negative likelihood ratio 0.18 (95% CI 0.10-0.32)

Positive likelihood ratio 5.51 (95% CI 3.11-9.77)

While the results are clearly inferior to those expected based on our internal validation, the the LTRC database is comprised of nodules with a very high pretest probability of malignancy considered suspicious enough to be surgically resected. Validation on a more similar dataset of incidentally identified lung nodules from the Vanderbilt University nodule registry is described below.

Year 3: Vanderbilt nodule cohort validation (see previous report for details)

Preliminary data from this validation study were reported in the previous report but have since been published in the European Respiratory Journal.

A total of 170 incidentally identified lung nodules from a well curated indeterminate pulmonary nodule registry database at Vanderbilt University, Nashville, TN, were used for independent external validation of the model described above. Cases were excluded due to missing slices, inability to segment or lack of sufficient information to calculate the Brock Score See flow chart below). The CT scans in DICOM format were transferred to Mayo Clinic Rochester, MN, for radiomic analysis.

**Comparison of Classifier with Brock Model**

The performance of Brock, a well validated nodule malignancy probability calculator widely used in clinical practice, developed from the Pan-Canadian Early Detection of Lung Cancer Study (Pan Can) was compared with our model in both the NLST LDCT and Vanderbilt incidentally detected nodule databases. Brock scores for 685 NLST nodules were calculated and compared with the prediction of our radiomic model by ROC analysis. In addition, comparative ROC analysis was performed on subsets of nodules classified based on pre-test malignancy probability as follows: low probability (Brock score <10%, N=350), intermediate probability (Brock score ≥10% but <60%, N=314) and high probability (Brock score ≥ 60%, N=21).

Validation data:

The Vanderbilt external validation set included 170 consecutive patients with incidentally identified IPNs (diameter 7-30 mm) enrolled into the Vanderbilt University pulmonary nodule registry. In the Vanderbilt University cohort, the mean diameter of the malignant nodules was larger than the benign nodules, 10.3 mm CI (9.4-11.3mm) versus 17.5 mm CI (16.2-17.8 mm), respectively ($p<0.001$) **(Figure 3)**. Confusion tables comparing the clinical/histological ground truth to the Brock model and the BRODERS classifier for the NLST and Vanderbilt datasets are shown in **Tables 1** and **2,** respectively. The distribution of malignancies and their BRODERS classifications at various Brock score categories are displayed in **Tables 3 and 4.**

Using the optimal cutoff of 0.478 identified via Youden's index, the sensitivity and specificity of the BRODERS classifier were 88.7% and 86.2% in the NLST screen-detected nodule cohort (n=685), respectively. For nodules with intermediate pre-test probability of malignancy (5-65%) by the Brock model (n=416) the Sensitivity was 91.9% and the Specificity was 71.6% using the same cutoff.

For the entire Vanderbilt incidental nodule dataset (n=170), the Sensitivity was 92.3%, the Specificity was 62.0%, the positive predictive value (PPV) was 73.7% and the negative predictive value (NPV) was 87.5%. For nodules with intermediate pre-test probability of malignancy by the Brock model (n=97), the Sensitivity was 94%, Specificity was 46%, the PPV was 78.4% and the NPV was 79.2%. The performance of the BRODERS classifier across different Brock-probability cut offs for the intermediate lung nodules are shown in **Table 5** and **6**.

The direct correlation between the Brock Model and the BRODERS classifier for the Vanderbilt University cohort are shown in **Figure 6**. **Figures 1 and 2** show the ROC comparing Brock model versus BRODERS for the entire NLST and Vanderbilt cohorts, and subsets of the cohort classified as low and intermediate pre-test malignancy risk. In both cohorts the AUC are significantly greater for the BRODERS model compared to the Brock model at all pre-test malignancy probabilities ($P<0.001$). The difference is most pronounced in the intermediate pre-test malignancy risk group. The benign resection rates based on the hypothetical application of the BRODERS classifier to the NLST and the Vanderbilt datasets are 12% and 26% for the entire cohorts and 10% and 22% for the Brock model intermediate probability nodules (5-65%).

**Table 1.** Truth tables comparing histology versus BRODERS classifier versus Brock model probability categories in the NLST cohort.

| Brock Model probability of malignancy | Clinical/ Histological Classification | BRODERS Benign | BRODERS Malignant |
|---|---|---|---|
| **Low < 5 %** (N = 257) | Benign (N = 204) | 192 | 12 |
| | Malignant (N = 53) | 17 | 36 |
| **Intermediate 5 ≤ to < 65** (N = 416) | Benign (N = 109) | 78 | 31 |
| | Malignant (N = 307) | 25 | 282 |
| **High ≥ 65** (N = 12) | Benign (N =0) | 0 | 0 |
| | Malignant (N = 12) | 0 | 12 |

**Table 2.** Truth tables comparing histology versus BRODERS classifier versus Brock model probability categories in the Vanderbilt cohort.

| Brock Model probability of malignancy | Clinical/ Histological Classification | BRODERS Benign | BRODERS Malignant |
|---|---|---|---|
| **Low < 5 %** (N = 42) | Benign (N = 38) | 30 | 8 |
| | Malignant (N = 4) | 2 | 2 |
| **Intermediate 5 ≤ to < 65** (N = 126) | Benign (N = 41) | 19 | 22 |
| | Malignant (N = 85) | 5 | 80 |
| **High ≥ 65** (N = 2) | Benign (N =0) | 0 | 0 |
| | Malignant (N = 2) | 0 | 2 |

**Table 3.** Types of malignancies in the NLST cohort and distribution across the Brock and BRODERS classification

| Histology | Brock < 5% | 5% <= Brock < 65% | Brock >= 65% | BRODERS Benign | BRODERS Malignant |
|---|---|---|---|---|---|
| **Adenocarcinoma** (N=268) | 34 | 224 | 10 | 39 | 229 |
| **Squamous cell carcinoma** (N=71) | 14 | 55 | 2 | 2 | 69 |
| **Large cell carcinoma** (N=18) | 3 | 15 | 0 | 1 | 17 |
| **Small Cell carcinoma** (N=11) | 2 | 9 | 0 | 0 | 11 |
| **Carcinoid** (N=4) | 0 | 4 | 0 | 0 | 4 |

**Table 4.** Types of malignancies in the Vanderbilt cohort and distribution across the Brock and BRODERS classification

| Histology | Brock < 5% | 5% <= Brock < 65% | Brock >= 65% | BRODERS Benign | BRODERS Malignant |
|---|---|---|---|---|---|
| **Adenocarcinoma** (N=60) | 2 | 57 | 1 | 3 | 57 |
| **Squamous cell carcinoma** (N=24) | 2 | 21 | 1 | 1 | 23 |
| **Large cell carcinoma** (N=3) | 0 | 3 | 0 | 0 | 3 |
| **Small Cell carcinoma** (N=3) | 0 | 3 | 0 | 1 | 2 |
| **Carcinoid** (N=1) | 0 | 1 | 0 | 1 | 0 |

**Table 5** Effect of different Brock cutoffs the intermediate probability group on BRODERS diagnostic performance in the NLST cohort.

| | Brock Score Cutoffs for the Intermediate Group | | | | | |
|---|---|---|---|---|---|---|
| | 5-60 | 5-65 | 5-70 | 10-60 | 10-65 | 10-70 |
| TN | 78 | 78 | 78 | 47 | 47 | 47 |
| FP | 31 | 31 | 31 | 23 | 23 | 23 |
| FN | 25 | 25 | 25 | 15 | 15 | 15 |
| TP | 273 | 282 | 287 | 229 | 238 | 243 |
| Sensitivity | 0.92 (0.92-0.92) | 0.9185 (0.918-0.919) | 0.92 (0.919-0.921) | 0.939 (0.938-0.94) | 0.941 (0.939-0.942) | 0.942 (0.941-0.943) |
| Specificity | 0.72 (0.71-0.72) | 0.7156 (0.713-0.718) | 0.7156 (0.713-0.718) | 0.671 (0.668-0.68) | 0.67 (0.667-0.675) | 0.671 (0.668-0.675) |
| PPV | 0.898 (0.896-0.899) | 0.9009 (0.899-0.902) | 0.9025 (0.902-0.904) | 0.909 (0.907-0.91) | 0.912 (0.911-0.913) | 0.914 (0.912-0.915) |
| NPV | 0.757 (0.754-0.759) | 0.757 (0.754-0.76) | 0.757 (0.755-0.76) | 0.758 (0.755-0.76) | 0.758 (0.755-0.761) | 0.758 (0.755-0.761) |
| BRR % | 10.2 | 9.9 | 9.75 | 9.13 | 8.8 | 8.65 |
| Brock-AUC 95% CI | 0.65 (0.6-0.69) | 0.66 (0.61-0.70) | 0.66 (0.62-0.71) | 0.61 (0.55-0.66) | 0.62 (0.57-0.68) | 0.63 (0.58-0.68) |
| Rad- AUC 95 % CI | 0.89 (0.86- 0.92) | 0.90 (0.86-0.92) | 0.9 (0.86-0.92) | 0.88 (0.84-0.92) | 0.89 (0.85-0.92) | 0.89 (0.85-0.92) |

TN – True Negatives; FP – False Positives; FN – False Negatives; TP – True Positives; PPV – Positive Predictive Value, NPV – Negative Predictive Value; BRR – Benign Resection Rate; Rad-AUC – Radiomics AUC

**Table 6** Effect of different Brock cutoffs the intermediate probability group on BRODERS diagnostic performance in the Vanderbilt University cohort.

| | Brock Score Cutoffs for the Intermediate Group | | | | | |
|---|---|---|---|---|---|---|
| | 5-60 | 5-65 | 5-70 | 10-60 | 10-65 | 10-70 |
| TN | 19 | 19 | 19 | 10 | 10 | 10 |
| FP | 22 | 22 | 22 | 13 | 13 | 13 |
| FN | 5 | 5 | 5 | 4 | 4 | 4 |
| TP | 78 | 80 | 81 | 70 | 72 | 73 |
| Sensitivity | 0.94 (0.938-0.941) | 0.94 (0.939-0.943) | 0.942 (0.94-0.943) | 0.946 (0.9440.948) | 0.947 (0.946-0.949) | 0.948 (0.946-0.95) |
| Specificity | 0.463 (0.458-0.468) | 0.463 (0.459-0.468) | 0.463 (0.456-0.468) | 0.435 (0.428-0.441) | 0.435 (0.428-0.441) | 0.435 (0.428-0.441) |
| PPV | 0.78 (0.777-0.783) | 0.784 (0.782-0.786) | 0.786 (0.784-0.789) | 0.843 (0.841-0.846) | 0.847 (0.845-0.85) | 0.849 (0.846-0.851) |
| NPV | 0.792 (0.786-0.797) | 0.792 (0.786-0.797) | 0.792 (0.786-0.797) | 0.714 (0.707-0.722) | 0.714 (0.707-0.722) | 0.714 (0.707-0.722) |
| BRR % | 22 | 21.57 | 21.36 | 15.66 | 15.29 | 15.11 |
| Brock-AUC 95% CI | 0.793 (0.711-0.861) | 0.798 (0.717-0.864) | 0.8 (0.72-0.866) | 0.743 (0.644-0.826) | 0.749 (0.652-0.831) | 0.753 (0.656-0.834) |
| Rad- AUC 95 % CI | 0.854 (0.779-0.911) | 0.856 (0.782-0.912) | 0.857 (0.784-0.913) | 0.85 (0.763-0.915) | 0.852 (0.767-0.916) | 0.854 (0.769-0.916) |

TN – True Negatives; FP – False Positives; FN – False Negatives; TP – True Positives; PPV – Positive Predictive Value, NPV – Negative Predictive Value; BRR – Benign Resection Rate; Rad-AUC – Radiomics AUC

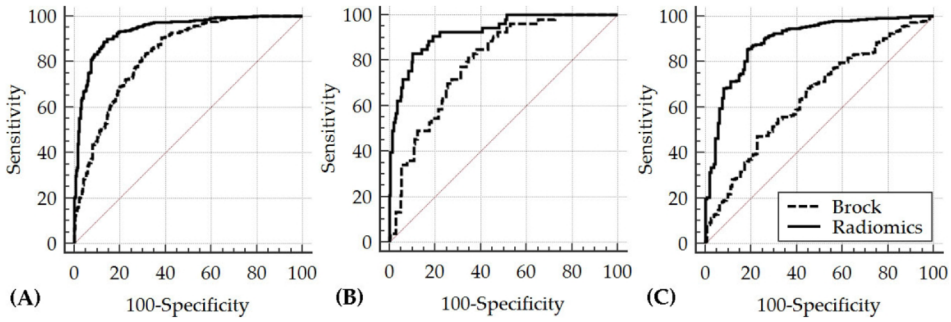


Figure 1. ROC for the NLST cohort comparing the Brock and Radiomics classifications. Panel (A) is for the entire cohort. AUC Brock 0.833 (95% CI = 0.803-0.860); AUC Radiomics 0.939 (0.918-0.955). Panel (B) is for the low risk (Brock score < 5%) group, AUC Brock 0.795 (0.74-0.842); AUC Radiomics 0.925 (0.886-0.954). Panel (C) is for the intermediate risk ( 5% ☐ Brock Score < 65%) group. AUC Brock 0.648 (0.599-0.694); AUC Radiomics 0.893 (0.859-0.922).

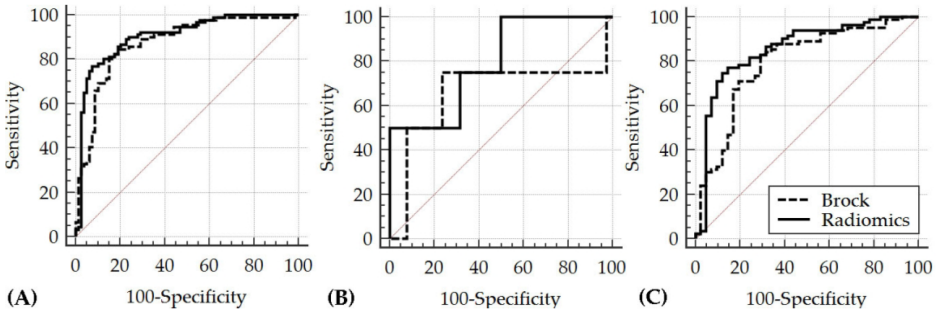


Figure 2. ROC for the Vanderbilt cohort comparing the Brock and Radiomics classifications. Panel (A) is for the entire cohort. AUC Brock 0.872 (95% CI = 0.812-0.918); AUC Radiomics 0.904 (0.849-0.943). Panel (B) is for the low risk (Brock score < 5%) group, AUC Brock 0.658 (0.496-0.797); AUC Radiomics 0.796 (0.644-0.904). Panel (C) is for the intermediate risk ( 5% ☐ Brock Score < 65) group. AUC Brock 0.798 (0.717-0.864); AUC Radiomics 0.856 (0.782-0.912).

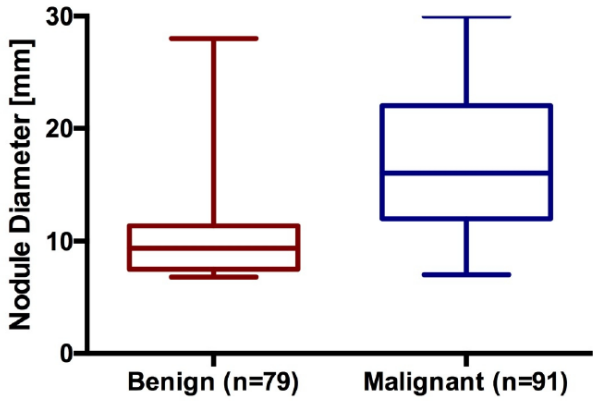


Figure 3

What opportunities for training and professional development has the project provided?

Nothing to report

How were the results disseminated to communities of interest?

Presentation at the DECAMP ½ Steering Committee meeting on June 15, 2020

Publication in the European Respiratory Journal (April 1, 2020)

What do you plan to do during the next reporting period to accomplish these goals?

As described in the initial proposal, we are still planning on validating our promising results on the DECAMP-1 dataset, which has now been transferred to Mayo Clinic Biomedical imaging Resource for image analysis. We have requested another no-cost extension for that purpose. A contract agreement between Vanderbilt University Medical Center and the American College of Radiology for use of clinical images was finalized. Once completed, results of radiomic analyses will then be transferred to Brown University for final analysis. We are planning on prospectively validating our model in an ongoing randomized controlled trial (Clinicaltrials.gov identifier NCT04250194) and are currently assembling a curated dataset of resected nodules to further validate our model. In the future, we plan on designing a prospective study to evaluate the impact of this novel radiomic biomarker on patient care in terms of clinical outcomes, morbidity, mortality, healthcare costs and pursue an FDA approval pathway.

## 4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

An estimated 1.5 million new lung nodules are identified via chest CT annually in the US. This figure is likely to increase with implementation of lung cancer screening for high-risk individuals, with individuals eligible for lung cancer screening estimated around 10 million in the US alone. In 2015, out of 8.6 million US individuals eligible for lung cancer screening as per NLST eligibility criteria, only 250,000 were screened, resulting in an estimated 750 lung cancer-related deaths averted, as opposed to the estimated 12,500 lives that could be saved with full implementation of lung cancer screening. One of the main obstacle to implementing lung cancer screening has been the large number of individuals with false positive screening CTs (approximately 40% in the NLST), is likely to result in unnecessary invasive diagnostic interventions with excessive morbidity, mortality, patient stress and healthcare expenses, particularly in individuals with lung nodules with intermediate pretest probability of malignancy.

We have previously demonstrated that volumetric CT-based quantitative characterization can risk-stratify lung nodules of the adenocarcinoma spectrum. This approach eliminates the intra- and inter-observer variability and subjectivity of CT image interpretation by trained radiologists. In addition, modern digital CT images include a large amount of valuable high-dimensional data not currently utilized to assist in diagnosis. We used to the NLST dataset to develop and internally validate a radiological multivariate model that include quantitative radiological features distinguishing malignant from benign CT-screen detected indeterminate pulmonary nodules. Initial validation of this model in independent cohorts has been promising and suggests that a

significant number of individuals with lung nodules could be spared additional non-invasive and invasive diagnostic interventions, mitigating the risk of unnecessary procedures associated with morbidity, mortality and healthcare costs. In addition, this tool leverages available data that are currently not exploited by clinicians and radiologists, obviating the need for further interventions, as required by other currently assessed biomarkers. This could lead to substantial improvement in lung nodule management, if available to a large audience of clinicians and radiologists as a software-based image analytical tool which could substantially reduce error and reduce the risk of unnecessary invasive and non-invasive procedures.

What was the impact on other disciplines?

Nothing to report

What was the impact on technology transfer?

Nothing to report

What was the impact on society beyond science and technology?

Our project, if successful could have a major impact on lung nodule management, by offering clinicians and radiologists reproducible tools to assist in the management of incidentally or screen-identified lung nodules, a major healthcare problem that affects Veteran and non-Veteran populations. Quantitative nodule analysis can be applied to existing CT scans obtained for screening or clinical indications and do not require additional testing beyond software application of image analytics. Our quantitative analytics tool could help standardize the management of lung nodules and lead to a substantial reduction in the unnecessary morbidity, mortality and healthcare costs.

## 5. CHANGES/PROBLEMS

Changes in approach and reasons for change:

There has not been a major change in our approach, except for the pursuit of additional validation sets given the considerable delays in accumulating enough cases in the DECAMP1 dataset to allow for enough power. None of the Department of Defense funds allocated to these analyses. We are now planning on analyzing the DECAMP-1 final results.

Actual or anticipated problems or delays and actions or plans to resolve them:

While this work was proposed to be completed last year, two unexpected major events prevented us from completing this work and motivated a request for another no-cost extension: (1) the COVID-19 pandemic, which placed a disproportionate burden on our division of pulmonary and critical medicine and slowed down our work considerably, and (2) the sudden and unexpected passing of Pierre Massion, MD, on April 4$^{th}$, 2021, who led the Massion Biomarker Laboratory at Vanderbilt University Medical Center, in which image analysis was to be performed, required us to modify our plans and transfer the images to the Biomedical Imaging Resource laboratory at Mayo Clinic, Rochester, MN, for further image segmentation and analysis. These datasets have now been successfully transferred.

This award was effective on September 30, 2015, but because of the relocation of the grant PI (Fabien Maldonado) from Mayo Clinic, Rochester, MN to Vanderbilt University, Nashville, TN, substantial delays were incurred from the need to establish subcontracts between the three partnering institutions (Mayo Clinic, Brown University and Vanderbilt University), which were eventually finalized in April 2016. This resulted in a significant delay for case selection and image transfer from the ACRIN and LSS core labs and our work on the development and optimization of discriminative radiological quantitative variables.

However, the variables were developed and optimized by the end of 2016 and both model 1 (radiological model) and model 2 (clinical-radiological model) were developed and internally validated using LASSO for variable penalization and selection and bootstrapping for internal validation. External validation, however, has been hampered by delays in recruitment in our planned validation dataset, the DECAMP1 dataset (PI: Dr. Avrum Spira). Accordingly, we have pursued additional validation cohorts and were able to validate our radiological model using the LTRC and Vanderbilt datasets.

Changes that had a significant impact on expenditures

Nothing to report.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, ad/or select agents

Nothing to report

## 6. PRODUCTS

Publications, conference papers, and presentations

1. Conference paper:
   Computed tomography-based radiomic classifier distinguishes malignant from benign nodules in the national screening trial
   18th World Conference on Lung Cancer
   October 15 - 18 2017 | Yokohama, Japan http://wclc2017.iaslc.org/
   Computed tomography-based radiomic classifier distinguishes malignant from benign nodules in the national screening trial
2. Journal publication:
   Peikert T, Duan F, Rajagopalan S, Karwoski RA, Clay R, Robb RA, Qin Z, Sicks J, Bartholmai BJ, Maldonado F. Novel high-resolution computed tomography-based radiomic classifier for screen-identified pulmonary nodules in the National Lung Screening Trial. PLoS One. 2018 May 14;13(5):e0196910.

   Maldonado F, Varghese C, Rajagopalan S, Duan F, Balar A, Lakhani D, Antic S, Massion PP, Johnson T, Karwoski,R, Robb R, Bartholmai B, Peikert T. Validation of the BRODERS classifier (Benign VS. aggressive nOdule Evaluation using Radiomic Stratification), a novel high-resolution computed tomography-based radiomic classifier

for indeterminate pulmonary nodules. European Respiratory Journal (published April 1, 2021)

Website(s) or other internet site(s):

Nothing to report

Technologies or techniques:

Novel CT-based quantitative analytics to distinguish benign from malignant nodules. How this novel analytical tool will be shared has not yet been determined.

Inventions, patent applications and/or licenses:

Nothing to report

Other products:

Nothing to report

## 7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?
No effort

Has there been a change in the active or other support of the PD/PI(s) or senior/key personnel since the last reporting period? See below.

**Fabien Maldonado, M.D.**

Nothing to report

**Tobias Peikert, M.D.**

Nothing to report

**Srinivasan Rajagopalan, Ph.D.**

Nothing to report

**Fenghai Duan, Ph.D.**

Nothing to report

What other organizations were involved as partners?
Nothing to report.

## 8. SPECIAL REPORTING REQUIREMENTS

Nothing to report.

## 9. APPENDICES

None.