AWARD NUMBER: W81XWH-15-1-0110

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

PRINCIPAL INVESTIGATOR: Fabien Maldonado

CONTRACTING ORGANIZATION: Vanderbilt University Medical Center
Nashville, TN

REPORT DATE: October 2017

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 2017	Annual	30Sep2016 - 29Sep2017

4. TITLE AND SUBTITLE

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

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-15-1-0110

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

Fabien Maldonado

E-Mail: Fabien.maldonado@vanderbilt.edu

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

Vanderbilt University Medical Center
Donald C. Brown
3319 West End Ave, STE. 910
Nashville, TN 37203

8. PERFORMING ORGANIZATION REPORT NUMBER

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

U.S. Army Medical Research and Materiel 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

Purpose: In the National Lung Screening Trial (NLST), indeterminate pulmonary nodules were detected in 40% of high-risk individuals screened by low dose high-resolution computed tomography (HRCT). However, 96% of these nodules were benign, indicating that false positive findings represent a major challenge for the clinical adoption of CT-based lung cancer screening. While current clinical-radiological risk prediction models are very valuable, optimization of the clinical management strategies for larger (≥ 7 mm) screen-detected nodules is needed to avoid unnecessary diagnostic interventions including futile thoracotomies. In this project, we explore the utility of a novel radiomics-based approach for the classification of screen-detected indeterminate nodules.

Material and methods: Independent quantitative variables assessing various radiologic nodule features such as sphericity, flatness, elongation, spiculation, lobulation and curvature were developed from the NLST dataset (using all 726 nodules > 7 mm; benign, n=318 and malignant, n=408). Multivariate analysis was performed using least absolute shrinkage and selection operator (LASSO) method for variable selection and regularization in order to enhance the prediction accuracy and interpretability of the multivariate model. To increase the stability of the modeling, LASSO was run 1,000 times and the variables that were selected in 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 (model 1: radiologic model). Relevant clinical variables (patient age and smoking history in pack-years) were then added to the model in an attempt to improve its diagnostic test characteristics (model 2: clinical-radiologic model).

Major findings: Eight radiologic features were selected by LASSO multivariate modeling out of 57 quantitative radiological variables considered for inclusion. These 8 features include variables capturing vertical location (centroid_Z), volume estimate (Min Enclosing Brick), flatness, texture analysis (SILA_Tex), surface complexity (Max_SI and Avg_SI), and estimates of surface curvature (Avg_PosMeanCurv and Min_MeanCurv), all with P<0.01. The optimism-corrected AUC for model 1 is 0.939. Our novel radiomic HRCT-based approach to non-invasive screen-detected nodule characterization appears extremely promising. We then added variables independently associated with an increased risk of lung cancer in our cohort (age and pack-years). The optimism-corrected

15. SUBJECT TERMS

lung adenocarcinoma – Radiomics – Lung cancer screening – chest computed tomography – biomarkers – lung nodules

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE		31	19b. TELEPHONE NUMBER *(include area code)*
U	U		UU		

Table of Contents

1. INTRODUCTION:

There are approximately 220,000 new lung cancers every year in the US, accounting for 160,000 deaths per year, more than colon, prostate and breast cancer combined. In 2011, the National Lung Screening Trial (NLST), a large randomized controlled trial on lung cancer screening, demonstrated a 20% relative reduction in lung cancer morality with annual low-dose chest computed tomography (LD-CT). These encouraging results have led to widespread endorsement of lung cancer screening, but at the cost of identifying many false-positive LD-CT. In the NLST, 40% of patients had identifiable lung nodules, 96% of which proved benign. In addition, there are approximately 20 million new chest CTs performed every year in the US, contributing to the identification a large reservoir of incidentally identified lung nodules, with an estimated 1.5 million new nodules detected every year. Currently, the detection of lung nodules leads to non-invasive and invasive studies to determine whether they are benign or malignant. Many patients with benign nodules are currently submitted to unnecessary procedures, increasing morbidity, mortality and healthcare costs. Novel tools to distinguish benign from malignant nodules are needed. We have previously demonstrated that volumetric CT-based quantitative imaging for lung adenocarcinoma characterization is useful in risk-stratifying these lesions, exploiting the wealth of data points available with modern CT imaging. In this project, we are attempting to use similar quantitative imaging metrics to help radiologists and clinicians determining the likelihood of malignancy based on radiologic (radiologic model) and combined clinical and radiologic characteristics (clinical-radiologic model). To do so, we used the available NLST dataset as a training set and will use the large ongoing prospective study Detection of Early lung Cancer Among Military Personnel Study 1 (DECAMP-1) for validation. This project will help to limit morbidity, mortality and healthcare costs associated with the management of incidentally or screen-identified pulmonary nodules.

2. KEYWORDS:

lung adenocarcinoma – Radiomics – Lung cancer screening – chest computed tomography – biomarkers – lung nodules.

3. ACCOMPLISHMENT:

3.1 What were the major goals of the project?

Aim 1 (first year of the grant): The first aim of this grant was to develop an imaging-based approach using volumetric analysis of screen-identified lung nodules, and a combined clinical-radiologic model 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.

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

The radiologic model was completed.

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

The clinical/radiologic model was completed, but addition of clinical variables did not contribute substantially to the diagnostic test performance of the model.

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).

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

Enrollment for the DECAMP1 study has been considerably delayed. Completion of enrollment in the study 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 DECAMP1 PI Dr. Avrum Spira). As of August 2017, DECAMP-1 study has accrued and adjudicated enough cases for validation (274 cases with 183 malignant and 91 confirmed benign nodules as of August 26, 2017). An application to access this dataset was completed (see supplement material) and recently submitted to the DECAMP biomarker committee for image transfer which is in preparation at this time.

In addition, we have now secured an alternative 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).

Finally, 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 were evaluated by expert radiologists and felt to be suspicious enough for malignancy to require surgical resection (see below).

3.2 What was accomplished under these goals:

Major activities:

Subject selection, summary of activities that occurred during the first year of the grant

The NLST was a randomized controlled trial conducted at 33 US centers, approved by the Institutional review boards at all participating centers. The study recruited asymptomatic high-

risk individuals from August 2002 through April 2004, aged 55 to 74 years, with a smoking history of at least 30 pack-years, having quit 15 years or less prior to randomization. Individuals were screened with either annual low-dose CT or chest X-ray for three years and followed through December 31, 2009. 26,722 individuals were randomized to the low-dose CT arm, and over 10,000 nodules (4-30 mm in longest diameter) were reported from at least one of the screening rounds.

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 screen-detected lung cancer cases: adenocarcinomas, squamous cell carcinomas, large cell carcinomas, small cell carcinomas and carcinoid tumors. 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 relative 1:1 fashion. Subsequently, it was decided that only one nodule per scan per participant would be analyzed, and, accordingly, CT with more than one nodule were analyzed as having only one nodule, in which case the largest nodule was selected. 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.

Screening HRCT data.

All NLST screening scans were low-dose scans with 2.5 mm collimation or less as pre-defined by strict NLST criteria, the details of which have been published elsewhere. The CT datasets were obtained from the Lung Screening Study core laboratory and transferred to a hard drive that was shipped to the investigators. The datasets from the American College of Radiology Imaging Network core laboratory were transferred initially via hard drive, then electronically to the investigators. Information on nodule location was available to the investigators in the NLST database and confirmed by one radiologist (B.J.B.) and two pulmonologists (F.M. and T.P.) using the CT obtained the closest in time to the diagnosis of malignant or benign lung nodules. Nodules were electronically tagged for segmentation and analysis. HRCT without visible nodules, nodules with borders indistinguishable from neighboring structures (e.g. mediastinum or pleura) and nodules without related clinical data were excluded.

Optimization and validation of nodule segmentation.

The lung nodules were segmented manually using the ANALYZE software (Biomedical Imaging Resource, Mayo Clinic, Rochester, MN). The location and the extent of each nodule was identified visually and a stack of two dimensional borders were traced out along the transverse orientation. A semi-automated region-growing approach based on the operator-specified bounding cube enclosing the nodule and a seed location within the nodule was used for initial segmentation. Manual editing was performed to remove, if needed, intruding structures like vessels and pleura. A parametric feature-based region growing technique based on the texture

classification of the voxels within the operator specified bounding cube was used as previously described.

Radiomic features:

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, as below:

1. Bulk metrics based on the global shape descriptors of the nodule.
2. Intensity metrics based on the CT Hounsfield units within the nodule.
3. Metrics capturing the spatial location of the nodule.
4. Nodule texture metrics based on the texture exemplar distributions within the nodule.
5. Surround texture metrics based on the parenchymal texture exemplar distributions within a region surrounding the nodule.
6. Metrics capturing the surface descriptors of the nodule.
7. Metrics capturing the distribution of the surface exemplars of the nodule.

Note that a considerable amount of work was performed during the 2nd year of the grant to refine and select the quantitative metrics initially present during the initial annual report.

Multivariate model (year 2 of the grant):

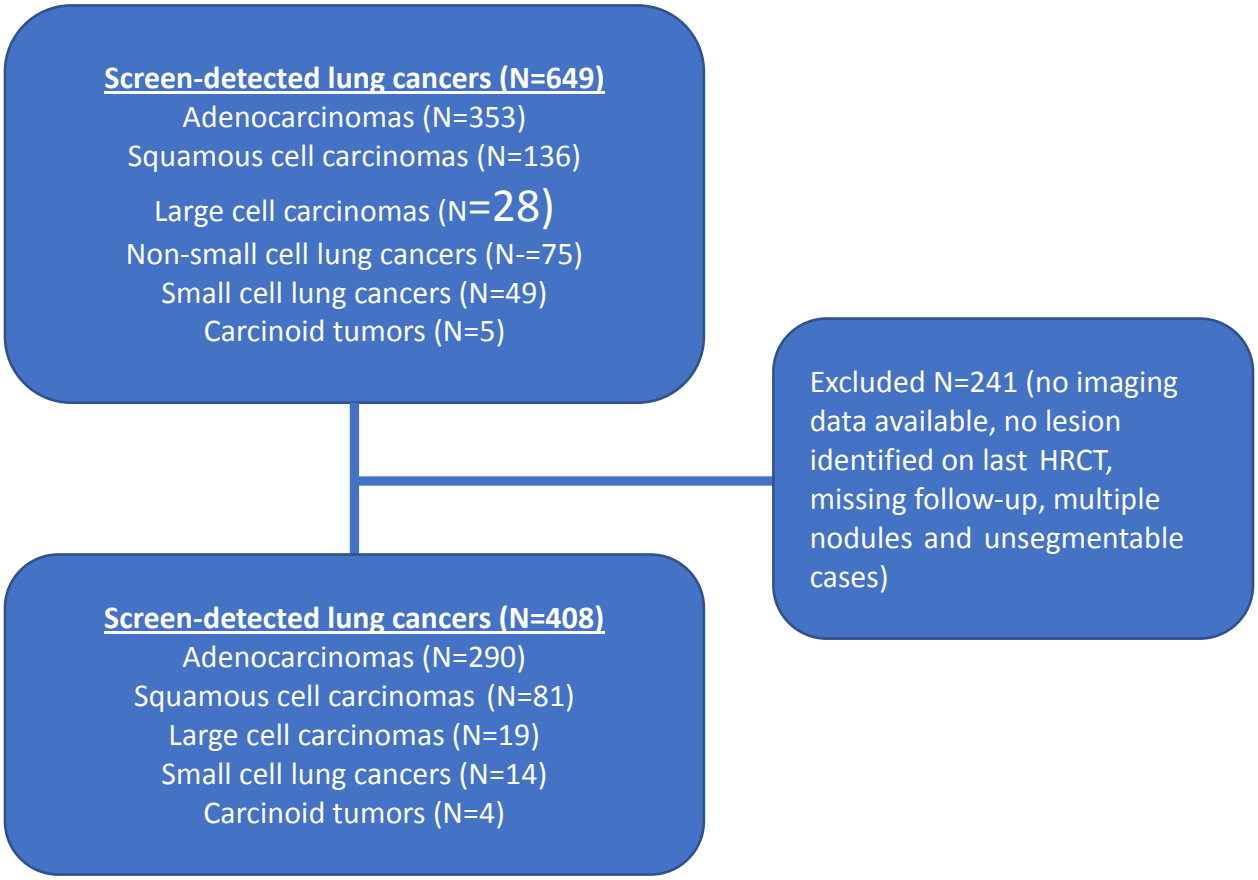
Quantitative methods were developed to characterize independent radiological variables assessing various radiologic nodule features such as sphericity, flatness, elongation, spiculation, lobulation and curvature using these nodules. Univariate analysis of the discriminatory power of each radiologic variable and receiver operative curve (ROC) analysis were performed for each variable and an area under the curve (AUC) calculated. Statistical significance was calculated and adjusted for multiple comparisons using Bonferroni correction. Spearman rank correlations between all pairs of variables were calculated and displayed via a heat map. 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.(19) The bootstrapping method was then applied for the internal validation, and the optimism-corrected AUC was reported for the final model.

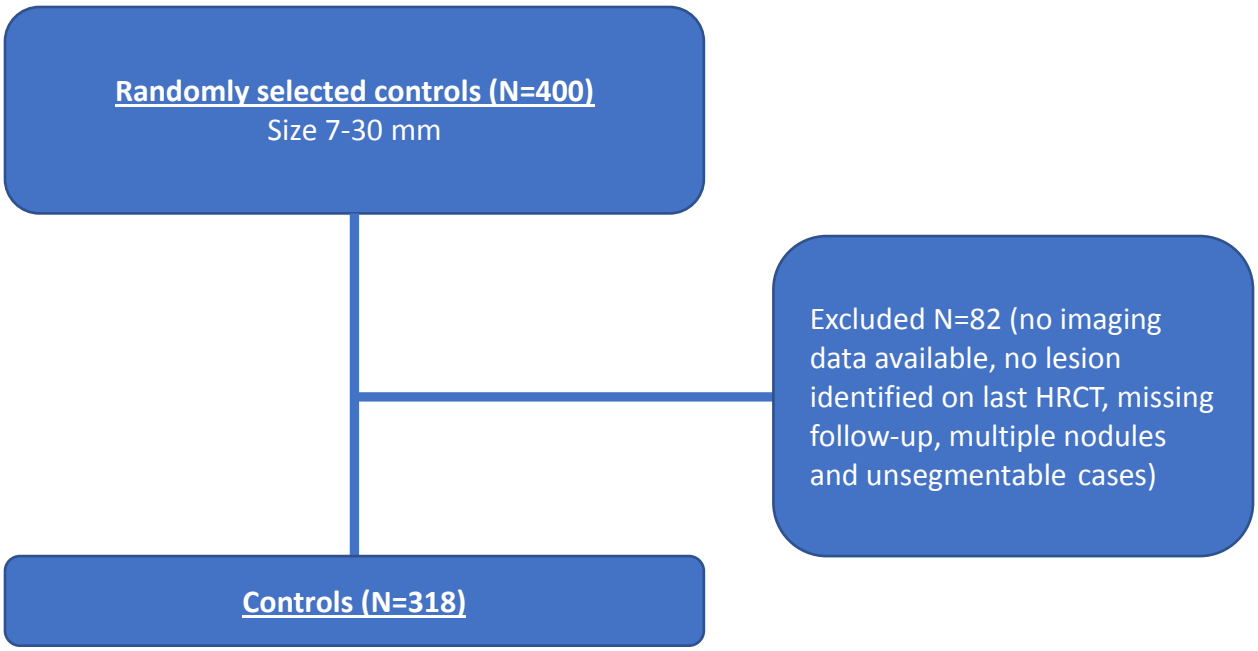
Results:

We reviewed 649 LDCT of cancers diagnosed in the screening arm of the NLST that included 353 adenocarcinomas, 136 squamous cell carcinomas, 28 large cell carcinomas, 75 non-small

cell carcinomas, 49 small cell carcinomas and 5 carcinoid tumors. 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.

Selection of cancer cases and controls, flowcharts:





The demographic and clinical characteristics of individuals included in the study are summarized below:

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.

In order to prevent overfitting of the model, we only considered quantitative imaging variables that were known *a priori* to be potentially associated with the benign or malignant nature of lung nodules. Quantitative methods were developed to characterize independent radiological variables assessing various radiologic nodule features such as 1. volume, 2. location, 3. surface characteristics (sphericity, flatness, elongation, spiculation, lobulation and curvature), 4. lung nodule texture features and 5. Lung texture analysis of the tumor-free surrounding lung, using 726 nodules identified from the NLST dataset (benign, n=318 and malignant, n=408) (see initial annual report).

AUC analysis across cancers and controls.

ID	Variables	Cancer_mean(SD)	Control_mean(SD)	AUC	P value
1	Centroid_x	154.78 (74.5)	142.21 (78.73)	0.56	0.02837
2	Centroid_y	143.95 (47.18)	151.84 (55.47)	0.47	0.03916
3	Centroid_Z	203.38 (60.1)	186.88 (65.91)	0.57	0.00052
4	Volume	3985.59 (13526.02)	344.48 (818.4)	0.9	0
5	SurfaceArea	1841.06 (3508.55)	344.12 (501.43)	0.87	0
6	Sphericity	0.51 (0.21)	0.6 (0.29)	0.58	1.00E-05
7	SphereFitFactor	6.82 (8.31)	5.28 (5.82)	0.58	0.00668
8	Radius_Estimated	7.61 (3.99)	3.59 (1.57)	0.9	0

9	Min.Enclosing.Brick_x	19.82 (12.12)	9.46 (5.51)	0.84	0
10	Min.Enclosing.Brick_y	19.63 (12.13)	10.11 (6.72)	0.82	0
11	Min.Enclosing.Brick	16.49 (14.51)	4.97 (2.65)	0.92	0
12	Max.Bricklength	24.08 (16.27)	11.31 (7.04)	0.84	0
13	Elongation	-0.25 (0.4)	-0.31 (0.47)	0.57	0.07783
14	Flatness	-0.56 (0.99)	-1.01 (1.05)	0.66	0
15	HU_mean	-209.18 (163.55)	-465.23 (201.91)	0.83	0
16	HU_var	614546.92 (3444392.14)	295011.7 (609422.64)	0.56	0.09419
17	HU_skew	-2.64 (10.09)	-2.39 (1.2)	0.57	0.66095
18	HU_kurt	133.91 (2032.65)	10.54 (10.04)	0.74	0
19	HU_entropy	7.89 (1.77)	6.76 (1.76)	0.82	0
20	Location	6.37 (3.42)	7.06 (3.16)	0.56	0.00558
21	SILA_Tex	122.91 (34.32)	58.62 (38.1)	0.88	0
22	Tex_Risk	2.17 (0.57)	1.36 (0.54)	0.82	0
23	Ves_.	1.88 (2.8)	0.75 (1.29)	0.74	0
24	Bgnd_.	9.49 (9.56)	9.59 (11.25)	0.52	0.89459
25	SILA_Fib	32.32 (17.84)	27.42 (22.96)	0.57	0.00136
26	SILA_Laa	35.54 (16.33)	32.69 (19.86)	0.55	0.03461
27	Num.Vertices	2711.4 (4745.67)	515.25 (697.45)	0.88	0
28	Num.Faces	5419.18 (9488.83)	1026.56 (1395.09)	0.88	0
29	WBE_2	1574.75 (3792.16)	480.61 (721.39)	0.75	0
30	WBE	2269.82 (6283.03)	802.67 (1116.04)	0.7	0
31	Min_MeanCurv	-0.92 (0.65)	-0.28 (0.46)	0.82	0
32	Max_MeanCurv	3.57 (2.44)	3.27 (1.82)	0.5	0.0694
33	Avg_PosMeanCurv	0.34 (0.11)	0.58 (0.2)	0.87	0
34	Skew_PosMeanCurv	2.89 (2.04)	2.01 (1.2)	0.66	0
35	Min_GCurv	-1.01 (0.87)	-0.87 (0.84)	0.58	0.03424
36	Max_GCurv	15.43 (30.41)	12.6 (21.14)	0.51	0.16811
37	Avg_PosGCurv	0.29 (0.29)	0.61 (0.52)	0.79	0
38	Skew_PosGCurv	7.57 (3.82)	4.66 (2.09)	0.78	0
39	Min_Sharp	0 (0)	0 (0)	0.79	0
40	Max_Sharp	38.99 (62.98)	22.44 (52.57)	0.59	0.00026
41	Avg_Sharp	0.59 (0.43)	1.01 (0.78)	0.71	0
42	Skew_Sharp	7.95 (7.45)	4.25 (3.53)	0.72	0
43	Min_Curved	0.01 (0.03)	0.07 (0.1)	0.82	0
44	Max_Curved	5.72 (4.21)	4.8 (3.05)	0.53	0.00131
45	Avg_Curved	0.58 (0.19)	0.96 (0.32)	0.87	0
46	Skew_Curved	2.87 (2.26)	1.79 (1.25)	0.69	0
47	Min_SI	-0.98 (0.01)	-0.98 (0.02)	0.63	0
48	Max_SI	0.98 (0.16)	0.55 (0.61)	0.82	0
49	Avg_SI	-0.29 (0.18)	-0.55 (0.13)	0.88	0
50	Skew_SI	1.63 (0.91)	1.72 (1.42)	0.54	0.3307

51	ICI	37.78 (118.81)	15.7 (21.56)	0.64	0
52	ECI	113.69 (284.16)	39.41 (57.05)	0.73	0
53	SILA_T	36.02 (11.24)	19.71 (12.61)	0.84	0
54	AvgCrv_T1	0.74 (0.23)	1.05 (0.32)	0.81	0
55	SkewCrv_T1	2.33 (1.73)	1.57 (1.04)	0.66	0
56	Avg_LocalSILA	27.65 (8.71)	15.3 (9.26)	0.84	0
57	Skew_LocalSila	0.71 (0.42)	0.49 (0.68)	0.6	0

Segmentation and reproducibility:

To assess the reproducibility and repeatability of the proposed segmentation, three operators (experienced radiologist, pulmonologist and image analyst) segmented multiple nodules (N = 266) from the NLST control cohort. The segmentation masks generated by the operators were compared pairwise using Dice Similarity Coefficient (DSC). The 95% confidence interval for the DSC between radiologist-pulmonologist, radiologist-image analyst and pulmonologist-image analyst was respectively 0.792-0.772, 0.785-0.804 and 0.835-0.857 (**see supplemental material**).

Radiomic features considered and selected

Intra-individual reproducibility: We used the **R**eference **I**mage **D**atabase to **E**valuate therapy **R**esponse (RIDER) dataset, a publicly available dataset of 31 paired CT scans obtained 15 minutes apart in the same individual using identical CT machine and acquisition protocol in patients with lung nodules to measure the reproducibility of the 57 initially selected radiomics variables. All 57 variables considered were found to be stable using all 3 paired tests (paired T, sign test and Wilcoxon).

Multivariate analysis

In order to select the optimal variables, 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 selected 57 quantitative imaging 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.

Model 1 – only radiomic features:

```
Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        -4.036621   1.661251  -2.430 0.015104 *
Centroid_Z          0.006502   0.001949   3.336 0.000849 ***
Min.Enclosing.Brick 0.206301   0.057399   3.594 0.000325 ***
SILA_Tex            0.023380   0.003821   6.118 9.48e-10 ***
Flatness            0.368149   0.221246   1.664 0.096116 .
Avg_PosMeanCurv    -1.292110   1.066489  -1.212 0.225683
Min_MeanCurv       -0.230528   0.367348  -0.628 0.530301
Max_SI              0.781022   0.411116   1.900 0.057464 .
Avg_SI              1.710727   1.765503   0.969 0.332558
```

AUC : 0.941

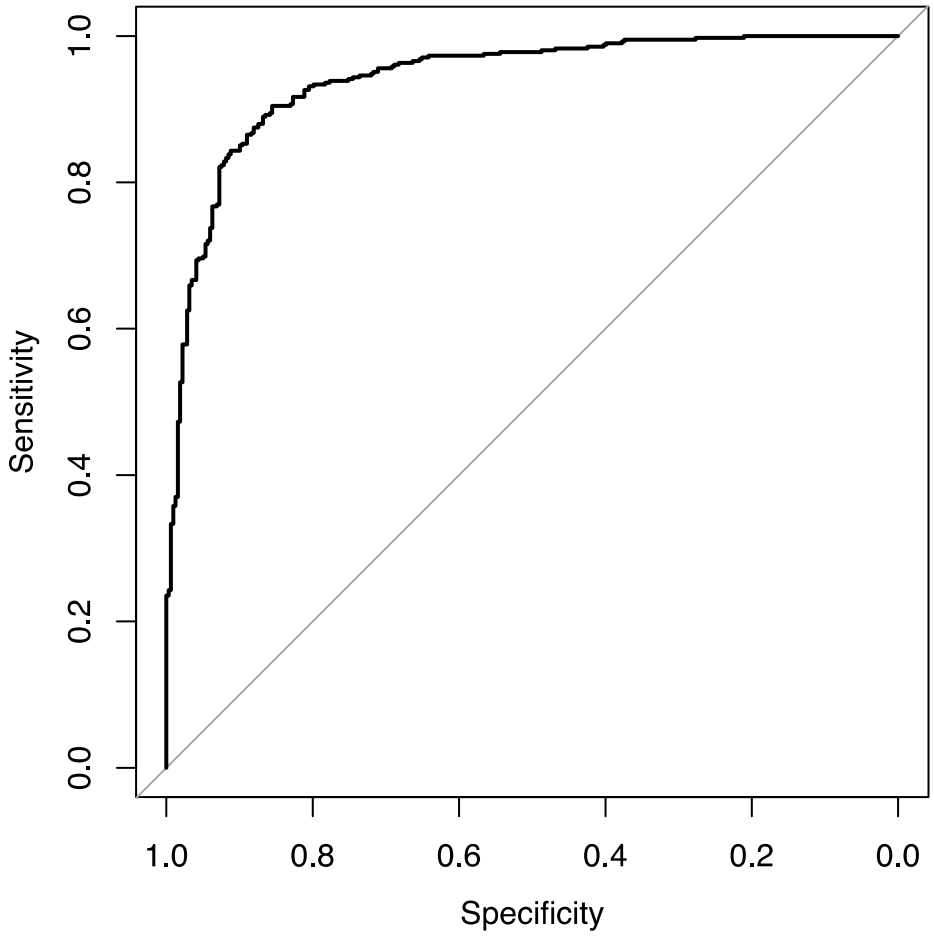
Optimism correction using bootstrap

```
Mean of Bootstrap AUC is 0.943
Mean of Test AUC is 0.941

The difference is 0.002
```

Optimism-corrected AUC for Model 1:

0.941 – 0.002 (optimism for model 1) = **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.

Model 2 – radiomic features + clinical variables

```
Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        -8.189458   2.390602  -3.426 0.000613 ***
Centroid_Z          0.005665   0.002095   2.704 0.006846 **
Min.Enclosing.Brick 0.178434   0.057463   3.105 0.001901 **
Flatness            0.390379   0.227995   1.712 0.086855 .
```

```
SILA_Tex             0.023527   0.004142   5.680 1.35e-08 ***
Min_MeanCurv        -0.332742   0.385821  -0.862 0.388454
Avg_PosMeanCurv     -1.425776   1.131270  -1.260 0.207550
Max_SI               0.663254   0.425328   1.559 0.118904
Avg_SI               1.759540   1.844368   0.954 0.340080
age                  0.063890   0.024851   2.571 0.010143 *
pkyr                 0.011214   0.005324   2.106 0.035171 *
```

AUC: 0.944

Optimism correction using bootstrap

```
Mean of Bootstrap AUC is 0.947
Mean of Test AUC is 0.944

The difference is 0.003
```

Optimism-corrected AUC for Model 2:

0.944 – 0.003 = **0.941**

Validation

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. Application to access this dataset was completed (see supplement material) and submitted to the DECAMP biomarker committee for image transfer.

In addition, we secured an alternative validation cohort from the lung nodule registry at Vanderbilt University Medical Center/Nashville Veterans Administration Tennessee Valley Healthcare system (primary investigator: Dr. Pierre Massion). All CT datasets have now been de-identified with corresponding clinical data recorded on a database currently unavailable to the investigators and password-protected at Vanderbilt University. We are now performing quality control on these CT datasets to ensure that they meet minimum criteria for radiomic analysis and tagging the nodules for analysis. We anticipate being done with this step by October 2017 with blinded analysis of all nodules (benign (n=100) and malignant (n=100)) by the end of November. This was approved by both Mayo Clinic and Vanderbilt University's respective institutional review boards.

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 were evaluated by expert radiologists and felt to be suspicious enough for malignancy to require surgical resection. Hence, this is a very different cohort than the cohort on which our radiologic model was derived (NLST), and we did not expect that it would perform as well.

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 nature of the LTRC database comprised of nodules with a very high pretest probability of malignancy make these results encouraging as we are in the process of validating these results on the more similar Vanderbilt and DECAMP1 database.

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

Nothing to report

3.4 How were the results disseminated to communities of interest?

Nothing to report

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

The development and internal validation of a radiological model using quantitative radiologic variables is now completed and extremely promising with an optimism-corrected area under the receiver operating curve of 0.939. Validation of this model on high-risk, resected suspicious lung nodules is also promising and suggests that our model may perform well in a validation cohort composed of individuals similar to the derivation cohort used (NLST). Due to considerable delays in recruitment in the DECAMP1 study, we have not yet been able to externally validate our results. Nonetheless, application to the DECAMP1 biomarker committee (Dr. Mark Lenburg) was submitted has enough benign and malignant nodules have now been adjudicated to allow for formal validation. This is based on the power calculation below:

This validation study requests a minimum of **274** cases who have been adjudicated in DECAMP 1, including **183** confirmed lung cancers and **91** confirmed benign disease. The classifier's performance will be assessed via calculating its discrimination and calibration. Discrimination measures the ability of the classifier's ability to differentiate lung cancers from the benign cases, which is commonly estimated through the ROC approach. The primary objective of this study is to determine if our classifier is significantly better than the model built up from the clinical and nodule features, such as age, smoking status, pack years, family history of lung cancer, nodule type and nodule location, etc. Using a two-sided z-test at a significance level of 0.05 and power of 90%, we can detect a difference between the AUC under the null hypothesis of **0.80** and an AUC under the alternative hypothesis of **0.878**, or between the AUC under the null hypothesis of **0.85** and an AUC under the alternative hypothesis of **0.918**. The sensitivity and specificity based on the optimal cutpoint(s) via Youden's index will also be validated.

Calibration is another important property of a classifier. We will perform the assessment through the calibration plotting (i.e., observed outcome versus predictions) and good-of-fit tests.

In addition, we have also secured an alternative 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 above). All CT datasets have now been de-identified with corresponding clinical data recorded on a database currently unavailable to the investigators and password-protected at Vanderbilt University. We are now performing quality control on these CT datasets to ensure that they meet minimum criteria for radiomic analysis and tagging the nodules for analysis.

While addition of clinical variables in our model 2 (clinical-radiological model) did not appear to provide superior performance of the model, it is possible that validation may be improved with model 2 and we plan on validating model as well and compare these two models.

4. IMPACT

1. 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, which is likely an underestimate given the ever-expanding use of HRCT in the US and in the world. This is also likely to increase markedly with implementation of lung cancer screening for high-risk individuals, with a number of individuals eligible for lung cancer screening estimated around 10 million in the US alone. Only approximately 10,000 individuals have been screened based on Medicare data as of May 2017. The large number of individual 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.

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. This invaluable unexploited resource can be leveraged by modern quantitative imaging methods. Radiomic approaches to lung nodule analysis consist of extracting reproducible and objective quantitative radiological variables from CT datasets, reducing large volumes of complex data into manageable and clinically relevant information. These quantitative imaging techniques have been proposed to facilitate the development of diagnostic and prognostic models in lung imaging, allowing for example the risk-stratification of lung adenocarcinomas, the classification of screen-or incidentally detected lung nodules and the characterization of lung cancer subtypes and tumor heterogeneity. 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. If this model is externally validated on a broad scale, it could lead to substantial improvement in lung nodule management, 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.

2. What was the impact on other disciplines?

   Nothing to report

3. What was the impact on technology transfer?

   Nothing to report

4. What was the impact on society beyond science and technology?

   Our project is not completed yet, but 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:**

1. Changes in approach and reasons for change:

   There hasn't been a major change in 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. Now that **274** cases have been adjudicated, including **183** confirmed lung cancers and **91** confirmed benign disease, we have the dataset and are awaiting image transfer. Next steps will include quality control, segmentation of nodules and radiomic analysis in a blinded fashion. This has resulted in a significant delay leading us to consider alternative validation datasets including CT datasets from the LTRC (lung tissue research consortium dataset) and the cohort from the lung nodule registry at Vanderbilt University Medical Center/Nashville Veterans Administration Tennessee Valley Healthcare system (primary investigator: Dr. Pierre Massion).

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

   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 dataset. This dataset, however, is significantly different than our derivation dataset in that all nodules were resected because of high suspicion of malignancy, explaining the decreased diagnostic test performance of our radiomic model. We are in the process of validating this model on another alternative dataset, the lung nodule registry at Vanderbilt University Medical Center/Nashville Veterans Administration Tennessee Valley Healthcare system (primary investigator: Dr. Pierre Massion). All images have been transferred and are currently undergoing quality control and analysis. We are also now awaiting image transfer from the DECAMP1 dataset (see above).

3. Changes that had a significant impact on expenditures

   Nothing to report.

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

   Nothing to report

6. PRODUCTS

1. Publications, conference papers, and presentations

   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/

   A journal manuscript is also in preparation at this time.

2. Website(s) or other internet site(s)

   Nothing to report

3. 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.

4. Inventions, patent applications and/or licenses

Nothing to report

5. Other products

Nothing to report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

Name: Tobias Peikert
Project Role: PD/PI
Research Identifier: N/A
Nearest Person Months: 1.0
Contribution to the Project: Mayo Clinic PI, administrative leadership at Mayo Clinic, review and selection of all benign NLST (nodules) training set and benign and malignant DECAMP nodules. Shared supervision of Dr. Rajagopalan and Ron Karwoski with Dr. Bartholmai. Participation in weekly team videoconferences.
Funding Support: No Changes

Name: Brian Bartholmai
Project Role: Co-Investigator
Research Identifier: 0000-0001-7834-6579
Nearest Person Months: 1.0
Contribution to the Project: Radiology leader and liaison to DECAMP team. Selection all technically appropriate DCAMP scans and selection of all benign NLST (nodules) training set and benign and malignant DECAMP nodules. Shared supervision of Dr. Rajagopalan and Ron Karwoski with Dr. Peikert. Participation in weekly team videoconferences.
Funding Support: NIH R01HL 125124-3 (Zhang)

Name: Srinivasan Rajagopalan
Project Role: Co-Investigator
Research Identifier: 0000-0003-3286-1529
Nearest Person Months: 6.0
Contribution to the Project: Image analysis and development of imaging variables. Participation in weekly meetings.
Funding Support: No Changes

Name: Fenghai Duan, PhD
Project Role: CSS subcontract PI
Researcher Identifier: 306213
Nearest person months worked: 1.2 CM
Contribution to project: This 2-year subcontract officially started in the beginning of 2016. Erin and Fenghai are working with the investigators to design the study, establish and support access to the clinical data and images of NLST and DECAMP, develop database linking clinical and radiological data for study analysis, develop analysis plan, and address methodological issues arising in the design and analysis, etc. To date, they have developed the plan and delivered the required clinical and radiological data to the Mayo Clinic.

Funding Support: None

Name: Erin Greco, MS
Project Role: Biostatistician
Researcher Identifier: 315034
Nearest person months worked: 1.38 CM
Contribution to project: This 2-year subcontract officially started in the beginning of 2016. Erin and Fenghai are working with the investigators to design the study, establish and support access to the clinical data and images of NLST and DECAMP, develop database linking clinical and radiological data for study analysis, develop analysis plan, and address methodological issues arising in the design and analysis, etc. To date, they have developed the plan and delivered the required clinical and radiological data to the Mayo Clinic.

Other Support Changes (since 2016 Annual Progress Report)

Maldonado, Fabien, M.D.
Ended: Centurion Medical Products (Maldonado); W81XWH-15-1-0110 (Maldonado)
New: W81XWH-17-1-0442 (Blackwell); VISE (

Duan, Fenghai, Ph.D.
Ended: U01 CA 190254 (Schnall); U01 CA 196408 (Dubinett); American College of Radiology (ACR) Schnall; Blue Earth Diagnostics (Duan)
New: None

Bartholmai, Brian, M.D.
Ended: LAM0110P03-15 (Bartholmai); LTRC (Bartholmai); W81XWH-15-1-0110 (Maldonado)
New: None

Srinivasan, Rajagopalan, Ph.D.
Nothing to Report

Tobias, Peikert, M.D.
Nothing to Report

8. SPECIAL REPORTING REQUIREMENTS

None

9. APPENDICES

See Attached documents

1. Avi Spira, MD, MSc (Boston Univeristy) Letter of Support
2. DECAMP Biomarker Committee Approval and Application
3. Publication Abstract

BU
Boston University School of Medicine

Boston University School of Medicine
Department of Medicine
Division of Computational Biomedicine

Medical Campus
72 East Concord Street, E-631
Boston, MA 02118-2308
T 617-414-6960
F 617-414-6999

August 28, 2014

Fabien Maldonado, M.D.
Assistant Professor Medicine
Pulmonary and Critical Care Medicine Mayo Clinic
200 First Street SW
Rochester, MN 55905

Dr. Maldonado:

As Principal Investigator of the Detection of Early Lung Cancer among Military Personnel 1 (DECAMP-1) I am delighted to express my enthusiastic support for your application "**Non-invasive characterization of indeterminate pulmonary nodules detected on chest high-resolution computed tomography**" for funding by the Department of Defense (DOD) Congressionally Directed Medical Research Program: Lung Cancer Research Program Idea Development Award (FOA: W81XWH-14-LCRP-IDA). As you know, the DECAMP-1 study is also funded by the DOD and the scientific aims or your project and DECAMP-1 are closely aligned—with the overall goal of validating biomarkers for pulmonary malignancy and tailoring care to optimize quality of life and outcomes for lung cancer. Specifically, the goal of DECAMP-1 is to investigate the use of various biomarkers during the evaluation of indeterminate pulmonary nodules (7 to 30 mm) detected in military personnel and veterans at high risk for lung cancer. The application of the quantitative CT technology proposed by you and your collaborators within the DECAMP cohort has great promise to advance the science of pulmonary nodule management and improve outcomes in this complex patient group by serving as an imaging biomarker that can be used in addition to the other molecular biomarkers obtained in the DECAMP trial to better detect and model disease.

Low-dose, volumetric high-resolution CT (HRCT) data for all DECAMP participants is being collected and stored at the American College of Radiology Imaging Network Core Laboratory. Dr. Bartholmai, a co-investigator for your project, currently serves as a scientific advisor to DECAMP regarding the imaging component of this multi-center project, and therefore the close collaboration between the DECAMP investigators and your proposed team is assured. The DECAMP participants are being enrolled at multiple sites throughout the United States including 7 Veterans Administration hospitals (VA Greater Los Angeles Health Care System, VA Eastern Colorado Health Care System, VA Boston Healthcare System, Philadelphia VA Medical Center, VA Pittsburgh Healthcare System, VA North Texas Health Care System, Nashville VA Medical Center), 4 designated military treatment facilities (the

Naval Medical Center San Diego, the San Antonio Military Medical Center, the Naval Medical Center in Portsmouth VA and the Walter Reed National Military Medical Center) and 3 academic centers (UCLA, Hospital of the Univ. of Pennsylvania and the Roswell Park Cancer Institute). The heterogeneity of this imaging data will allow for validation of the quantitative CT analysis tools that should thus have real-world application.

Pulmonary nodules discovered in the DECAMP-1 trial will be managed based on standard international guidelines and clinical standard of care, including serial imaging, advanced techniques such as PET/CT and tissue biopsies with the overall goal of improving survival and quality of life for the participants in the screening trial. We are committed to provide imaging (HRCT data), clinical baseline, and clinical outcome data for your proposed project. Thus far we have recruited 125 of the planned 500 patients with a diagnosis of an indeterminate pulmonary nodule. We project to complete patient enrollment by December 2015.

We are very enthusiastic about continuing and expanding already existing scientific collaborations between our research groups. In addition to serving as the biostatistician for your project, Dr. Fenghai Duan is also the biostatistician for the DECAMP-1 study.

In summary I am very much looking forward to working with you and your team on this extremely promising project. I am excited that this new approach will significantly advance our clinical approach to patients presenting with indeterminate pulmonary nodules.

Sincerely,

Avi Spira, M.D., MSc
Professor of Medicine, Pathology, & Bioinformatics
Chief, Section of Computational Biomedicine
Department of Medicine
Boston University, School of Medicine

From:	Brewer, Katrina A
To:	Brewer, Katrina A
Subject:	FW: Radiomic model for indeterminate lung nodules
Date:	Monday, October 09, 2017 3:41:22 PM
Attachments:	decamp_biomarker_validation_application_fd.docx ATT00001.html

---

Begin forwarded message:

From: Elizabeth Moses <emoses@bu.edu>
Subject: Re: Radiomic model for indeterminate lung nodules
Date: October 6, 2017 at 5:24:37 PM CDT
To: "Maldonado, Fabien" <fabien.maldonado@vanderbilt.edu>
Cc: "Lenburg, Marc E" <mlenburg@bu.edu>, "Bartholmai, Brian J.," <Bartholmai.Brian@mayo.edu>, "Peikert, Tobias," <Peikert.Tobias@mayo.edu>, "Spira, Avrum" <aspira@bu.edu>, Fenghai Duan <fduan@stat.brown.edu>, "Bauza, Joseph" <jbauza@acr.org>

Hello Fabian,

Great news, the DECAMP biomarker committee has approved your request for access to a minimum of 274 CT images from DECAMP 1!

I am cc'ing Fenghai as well as Joe Bauza from ACRIN who can work with you to obtain access to these images. I am also attaching your application for everyone's reference.

Please let me know if I can be of assistance to anyone during the process, and I hope everyone has a nice weekend!

Best,
Liz
……

Elizabeth S. Moses, Ph.D. | Scientific Program Manager, DECAMP

Boston University School of Medicine
Section of Computational Biomedicine
Spira/Lenburg Lab
72 E. Concord St, Evans Building, 6th Floor | Boston, MA 02118

Investigator Contact Information

Fabien Maldonado, MD
Associate Professor of Medicine and Thoracic Surgery
Division of Allergy, Pulmonary and Critical Care Medicine
1161 21st Avenue South
T-1218 Medical Center North
Nashville, TN 37232
Fabien.maldonado@vanderbilt.edu

What is requested

A minimum of 274 (or more) CT datasets and clinical data of patients with adjudicated pulmonary nodules 7-30 mm enrolled in DECAMP-1 for validation of a CT-based radiomic classifier for indeterminate lung nodules.

Minimal sample amount required

All adjudicated lung cancers and benign disease (a minimum of 274) in DECAMP 1

Expected length of study

6 months

IRB Approval (yes/no/pending)

Yes

Funding for proposed studies

Lung Cancer Research Program, Innovative idea
X81XWH-15-1-0110 (Maldonado)
Department of Defense
"Non-Invasive Characterization of Indeterminate Pulmonary Nodules Detected on Chest High-Resolution Computed Tomography"
PI: Fabien Maldonado

Intellectual Property Status of biomarker

Clinical Question: Clearly state the clinical question/need that the biomarker seeks to address. How would access to samples and data from the DECAMP studies to expedite addressing the intended clinical question?

We are planning on validating a radiomic classifier for indeterminate pulmonary nodules developed using the National Lung Screening Trial Database. Our internally validated multivariate model includes 8 quantitative radiologic variables and has an optimism-corrected AUC of 0.939.

As proposed in this DOD-funded project, we would like to validate this model using indeterminate nodules from the DECAMP1 dataset.

Background and Significance: Clearly state the scientific rationale of the proposal for using the requested DECAMP samples and data. Describe your biomarker/platform and how you came upon its discovery/development.

Lung cancer accounts for more cancer-related deaths in the US than colon, prostate and breast cancer combined, approximately 160,000 deaths per year. In 2011, a large randomized controlled trial, the National Lung Screening Trial (NLST) demonstrated a 20% relative reduction in lung cancer morality with annual low-dose chest computed tomography (LD-CT). These encouraging results have led to widespread endorsement of lung cancer screening, but at the cost of identifying many false-positive LD-CT. In the NLST, 40% of patients had identifiable lung nodules, 96% of which proved benign. The ever-expanding use of chest CT in the US (estimated 20 million/year) is contributing to the identification of an estimated 1.5 million new nodules every year. Novel tools to distinguish benign from malignant nodules are urgently needed. Our team has previously demonstrated that volumetric CT-based quantitative characterization of lung nodules belonging to the adenocarcinoma spectrum is useful in risk-stratifying these lesions, leveraging the wealth of unexploited data points available with modern CT imaging. In this project, we used similar quantitative imaging metrics to assist radiologists and clinicians in determining the likelihood of malignant nodule based on radiologic characteristics. To do so, we used the available NLST dataset as a training set and internally validated the model. We are now seeking access to the large ongoing prospective study Detection of Early lung Cancer Among Military Personnel Study 1 (DECAMP-1) in order to validate this promising multivariate model. This project, if successful, will help to limit morbidity, mortality and healthcare costs associated with the management of incidentally or screen-identified pulmonary nodules.

Preliminary Data & Methods: Provide sufficient information describing how experiments were performed, details on the cohorts that have been studied, and presentation of data in terms of analytic validity, specificity, sensitivity, and variance of your measurements. Explicit description of your studies will facilitate review considerations. Figures and other supporting documentation can be appended to your proposal.

Methods:

Subject selection

The NLST was a randomized controlled trial conducted at 33 US centers, approved by the Institutional review boards at all centers. The study recruited asymptomatic high-risk individuals from August 2002 through April 2004, aged 55 to 74 years, with a smoking history of at least 30 pack-years, having quit 15 years or less prior to randomization. Individuals were screened with either annual low-dose CT or chest X-ray for three years and followed through December 31, 2009. 26,722 individuals were randomized to the low-dose CT arm, and over 10,000 nodules (4-30 mm in longest diameter) were reported from at least one of the screening rounds.

Participants for the present study 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 screen-detected lung cancer cases: adenocarcinomas, squamous cell carcinomas, large cell carcinomas, small cell carcinomas and carcinoid tumors. 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 relative 1:1 fashion. Subsequently, it was decided that only one nodule per scan per participant would be analyzed, and, accordingly, CT with more than one nodule were analyzed as having only one nodule, in which case the largest nodule was selected. 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.

Screening HRCT data.

All NLST screening scans were low-dose scans with 2.5 mm collimation or less as pre-defined by strict NLST criteria, the details of which have been published elsewhere. The CT datasets were obtained from the Lung Screening Study core laboratory and transferred to a hard drive that was shipped to the investigators. The datasets from the American College of Radiology Imaging Network core laboratory were transferred initially via hard drive, then electronically to the investigators. Information on nodule location was available to the investigators in the NLST database and confirmed by one radiologist (B.J.B.) and two pulmonologists (F.M. and T.P.) using the CT obtained the closest in time to the diagnosis of malignant or benign lung nodules. Nodules were electronically tagged for segmentation and analysis. HRCT without visible nodules, nodules with

borders indistinguishable from neighboring structures (e.g. mediastinum or pleura) and nodules without related clinical data were excluded.

Optimization and validation of nodule segmentation.

The lung nodules were segmented manually using the ANALYZE software (Biomedical Imaging Resource, Mayo Clinic, Rochester, MN). The location and the extent of each nodule was identified visually and a stack of two dimensional borders were traced out along the transverse orientation. A semi-automated region-growing approach based on the operator-specified bounding cube enclosing the nodule and a seed location within the nodule was used for initial segmentation. Manual editing was performed to remove, if needed, intruding structures like vessels and pleura. A parametric feature-based region growing technique based on the texture classification of the voxels within the operator specified bounding cube was used as previously described.

Radiomic features:

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, as below:

1. Bulk metrics based on the global shape descriptors of the nodule.
2. Intensity metrics based on the CT Hounsfield units within the nodule.
3. Metrics capturing the spatial location of the nodule.
4. Nodule texture metrics based on the texture exemplar distributions within the nodule.
5. Surround texture metrics based on the parenchymal texture exemplar distributions within a region surrounding the nodule.
6. Metrics capturing the surface descriptors of the nodule.
7. Metrics capturing the distribution of the surface exemplars of the nodule.

Multivariate model:

Quantitative methods were developed to characterize independent radiological variables assessing various radiologic nodule features such as sphericity, flatness, elongation, spiculation, lobulation and curvature using these nodules. Univariate analysis of the discriminatory power of each radiologic variable and receiver operative curve (ROC) analysis were performed for each variable and an area under the curve (AUC) calculated. Statistical significance was calculated and adjusted for multiple comparisons using Bonferroni correction. Spearman rank correlations between all pairs of variables were calculated and displayed via a heat map. 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.(19) The bootstrapping method was then applied for the internal validation, and the optimism-corrected AUC was reported for the final model.

Results:

Study participants:

We reviewed 649 LDCT of cancers diagnosed in the screening arm of the NLST that included 353 adenocarcinomas, 136 squamous cell carcinomas, 28 large cell carcinomas, 75 non-small cell carcinomas, 49 small cell carcinomas and 5 carcinoid tumors. 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 in **Table 1**.

In order to prevent overfitting of the model, we only considered quantitative imaging variables that were known *a priori* to be potentially associated with the benign or malignant nature of lung nodules. Quantitative methods were developed to characterize independent radiological variables assessing various radiologic nodule features such as 1. volume, 2. location, 3. surface characteristics (sphericity, flatness, elongation, spiculation, lobulation and curvature), 4. lung nodule texture features and 5. Lung texture analysis of the tumor-free surrounding lung, using 726 nodules identified from the NLST dataset (benign, n=318 and malignant, n=408).

Segmentation and reproducibility:

To assess the reproducibility and repeatability of the proposed segmentation, three operators (experienced radiologist, pulmonologist and image analyst) segmented multiple nodules (N = 266) from the NLST control cohort. The segmentation masks generated by the operators were compared pairwise using Dice Similarity Coefficient (DSC; **Figure 2**). The 95% CI for the DSC between radiologist-pulmonologist, radiologist-image analyst and pulmonologist-image analyst was respectively 0.792-0.772, 0.785-0.804 and 0.835-0.857 (**see supplemental material**).

Radiomic features considered and selected

Intra-individual reproducibility: We used the **R**eference **I**mage **D**atabase to **E**valuate therapy **R**esponse (RIDER) dataset, a publicly available dataset of 31 paired CT scans obtained 15 minutes apart in the same individual using identical CT machine and acquisition protocol in patients with lung nodules to measure the reproducibility of the 57 initially selected radiomics variables. All 57 variables considered were found to be stable using all 3 paired tests (paired T, sign test and Wilcoxon).

Multivariate analysis

In order to select the optimal variables, 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 quantitative imaging 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.

Data Analysis Plan: Provide adequate detail concerning how statistical analysis of your data generated from the Reference Set(s) samples will be performed and a justification that the requested References Set(s) is/are large enough to demonstrate the utility of the biomarker. Describe the statistical resources at your disposal.

This validation study requests a minimum of **274** cases who have been adjudicated in DECAMP 1, including **183** confirmed lung cancers and **91** confirmed benign disease. The classifier's performance will be assessed via calculating its discrimination and calibration. Discrimination measures the ability of the classifier's ability to differentiate lung cancers from the benign cases, which is commonly estimated through the ROC approach. The primary objective of this study is to determine if our classifier is significantly better than the model built up from the clinical and nodule features, such as age, smoking status, pack years, family history of lung cancer,

nodule type and nodule location, etc. Using a two-sided z-test at a significance level of 0.05 and power of 90%, we can detect a difference between the AUC under the null hypothesis of **0.80** and an AUC under the alternative hypothesis of **0.878**, or between the AUC under the null hypothesis of **0.85** and an AUC under the alternative hypothesis of **0.918**. The sensitivity and specificity based on the optimal cutpoint(s) via Youden's index will also be validated. Calibration is another important property of a classifier. We will perform the assessment through the calibration plotting (i.e., observed outcome versus predictions) and good-of-fit tests.

Collaboration: In this section state your willingness to deposit all primary data obtained using DECAMP samples with the DECAMP Data Management and Coordinating Center (DMCC).

We would be willing to deposit all primary data using DCAMP samples with the DECAMP Data Management and Coordinating Center (DMCC).

Future Plans: If the biomarker is found to have promising performance characteristics, the DECAMP Consortium might be interested in working with you to proceed to Phase II clinical validations.

- Are you amenable to working within the collaborative framework of DECAMP in proceeding to Phase II studies?

  Yes

- Do you have other resources where validation studies can be accomplished? If so, describe clearly other resources at your disposal and how they could be used to complete a larger Phase II validation study.

  Vanderbilt University Medical Center and mayo Clinic lung nodule cohorts and lung cancer screening registries.

- If deemed beneficial, will you be amenable to including your biomarker into a larger panel of biomarkers for Phase II validation?

  Yes.

Please do not exceed 4 pages

Abstract 10244

Computed tomography-based radiomic classifier distinguishes malignant from bengin pulmonary nodules in the National Lung Screening Trial

Type: Peer Review

Topic: 13. Radiology/Staging/Screening

Authors: Tobias Peikert[1], Fenghai Duan[2], Srinivasan Rajagopalan[1], Ronald Karwoski[1], Ziling Qin[2], JoRean Sicks[2], Ryan Clay[1], Richard Robb[1], Brian Bartholmai[1], Fabien Maldonado[3]; [1]Mayo Clinic, Rochester/United States of America, [2]Brown University, Providence/United States of America, [3]Vanderbilt University, Nashville, Tn, MN/United States of America

Background

In the National Lung Screening Trial (NLST), indeterminate pulmonary nodules were detected in 40% of high-risk individuals screened by low dose high-resolution computed tomography (HRCT). However 96% of these nodules were benign indicating that overdiagnosis represents a major challenge for the clinical implantation of CT based lung cancer screening. While current clinical-radiological risk prediction models are very valuable, optimization of the clinical management of larger (≥ 7 mm) screen-detected nodules to avoid unnecessary diagnostic interventions including futile thoracotomies better strategies are needed. Herein we demonstrate the potential value of a novel radiomics based approach for the classification of screen-detected indeterminate nodules.

Method

Independent quantitative variables assessing various radiologic nodule features such as sphericity, flatness, elongation, spiculation, lobulation and curvature, using 726 nodules (all ≥ 7 mm) were developed from the NLST dataset (benign, n=318 and malignant, n=408). Multivariate analysis was performed using least absolute shrinkage and selection operator (LASSO) method for variable selection and regularization in order to enhance the prediction accuracy and interpretability of the multivariate model. To increase the stability of the modeling, LASSO was run 1,000 times and the variables that were selected in 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.

Result

Eight radiologic features were selected by LASSO multivariate modeling out of 57 quantitative radiological variables considered for inclusion. These 8 features include variables capturing vertical location (centroid_Z), volume estimate (Min Enclosing Brick), flatness, texture analysis (SILA_Tex), surface complexity (Max_SI and Avg_SI), and estimates of surface curvature (Avg_PosMeanCurv and Min_MeanCurv), all with $P<0.01$. The optimism-corrected AUC is 0.939.

Conclusion

Conclusion Our novel radiomic HRCT-based approach to non-invasive screen-detected nodule characterization appears extremely promising. Independent external validation is needed.