Big Data and Vulnerable Populations – Addressing the Gap

David Yankelevitz

Financial and Research Disclosures:

- Dr. David Yankelevitz is a named inventor on a number of patents and patent applications relating to the evaluation of diseases of the chest including measurement of nodules. Some of these, which are owned by Cornell Research Foundation (CRF) are non-exclusively licensed to General Electric. As an inventor of these patents, Dr. Yankelevitz is entitled to a share of any compensation which CRF may receive from its commercialization of these patents
- A shareholder in Accumetra LLC
- Medical Advisory Board Carestream Health
- Advisory Panel Pfizer, Genentech, AstraZeneca, LungLifeAI

Panelists

- Anthony P. Reeves, PhD, School of Electrical and Computer Engineering,

Cornell University

- Melinda Aldrich, PhD, MPH**,** Department of Medicine,

Vanderbilt University Medical Center

- Emanuela Taioli, MD, PhD, Institute for Translational Epidemiology,

Icahn School of Medicine at Mount Sinai

- Heather Pierce, JD, MPH, Center for Health Justice,

Association of American Medical Colleges

- Erik Lium, PhD, *Mount Sinai Innovation Partners*

Mount Sinai Health System

Pathway to development of a useful database

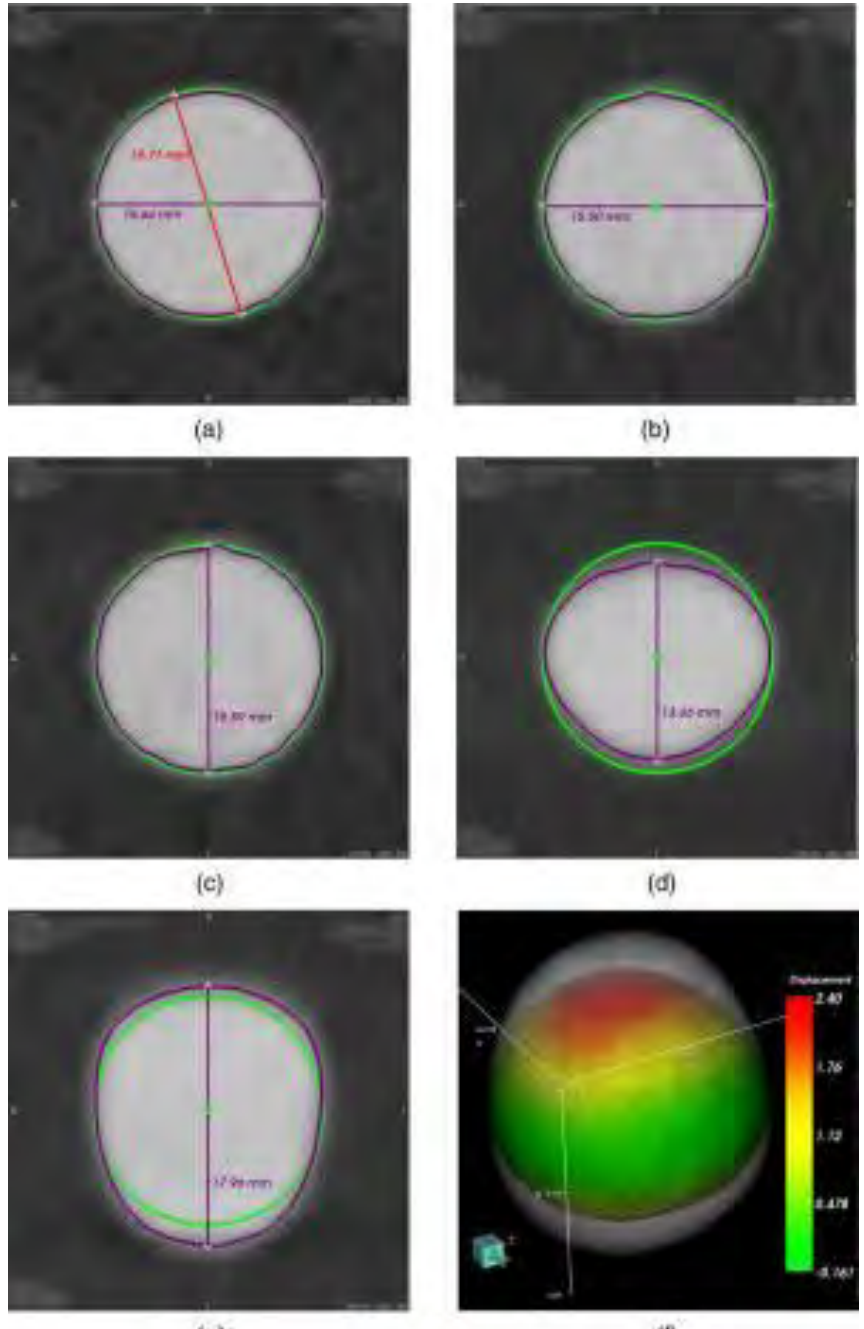
- LDCT in the context of lung cancer screening
- Need to define the purpose of images (ie. advanced image processing, quantitation, detection, diagnosis)
  - Lung cancer, emphysema, CAC, osteoporosis, other health measures
- Necessary quality for defined purpose

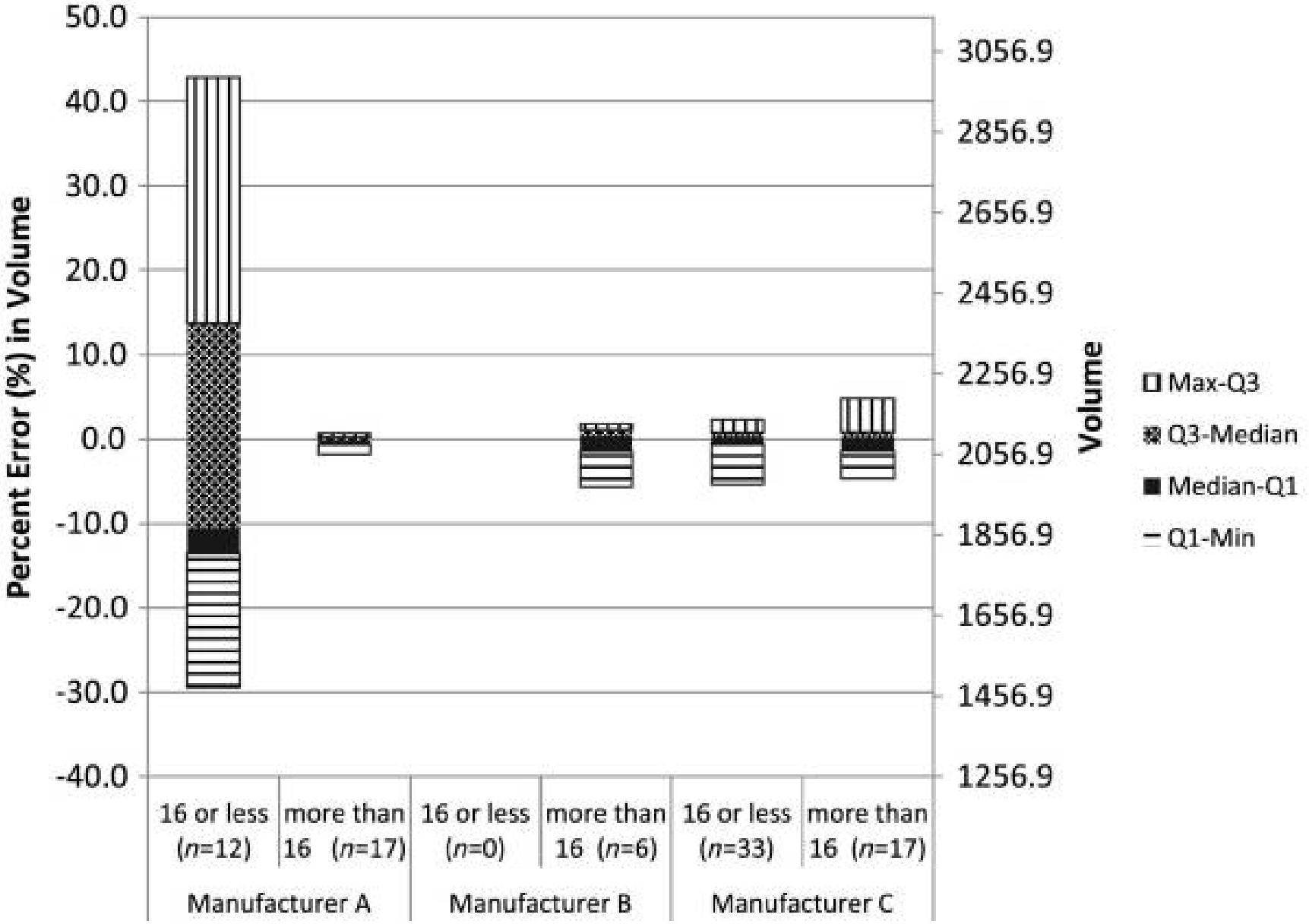
Journal of Medical Imaging 3(3), 035505 (Jul–Sep 2016)

Tumor volume measurement error using computed tomography imaging in a phase II clinical trial in lung cancer

Claudia I. Henschke,[a,b,*] David F. Yankelevitz,[a] Rowena Yip,[a] Venice Archer,[c] Gudrun Zahlmann,[d] Karthik Krishnan,[e] Brian Helba,[e] and Ricardo Avila[f]

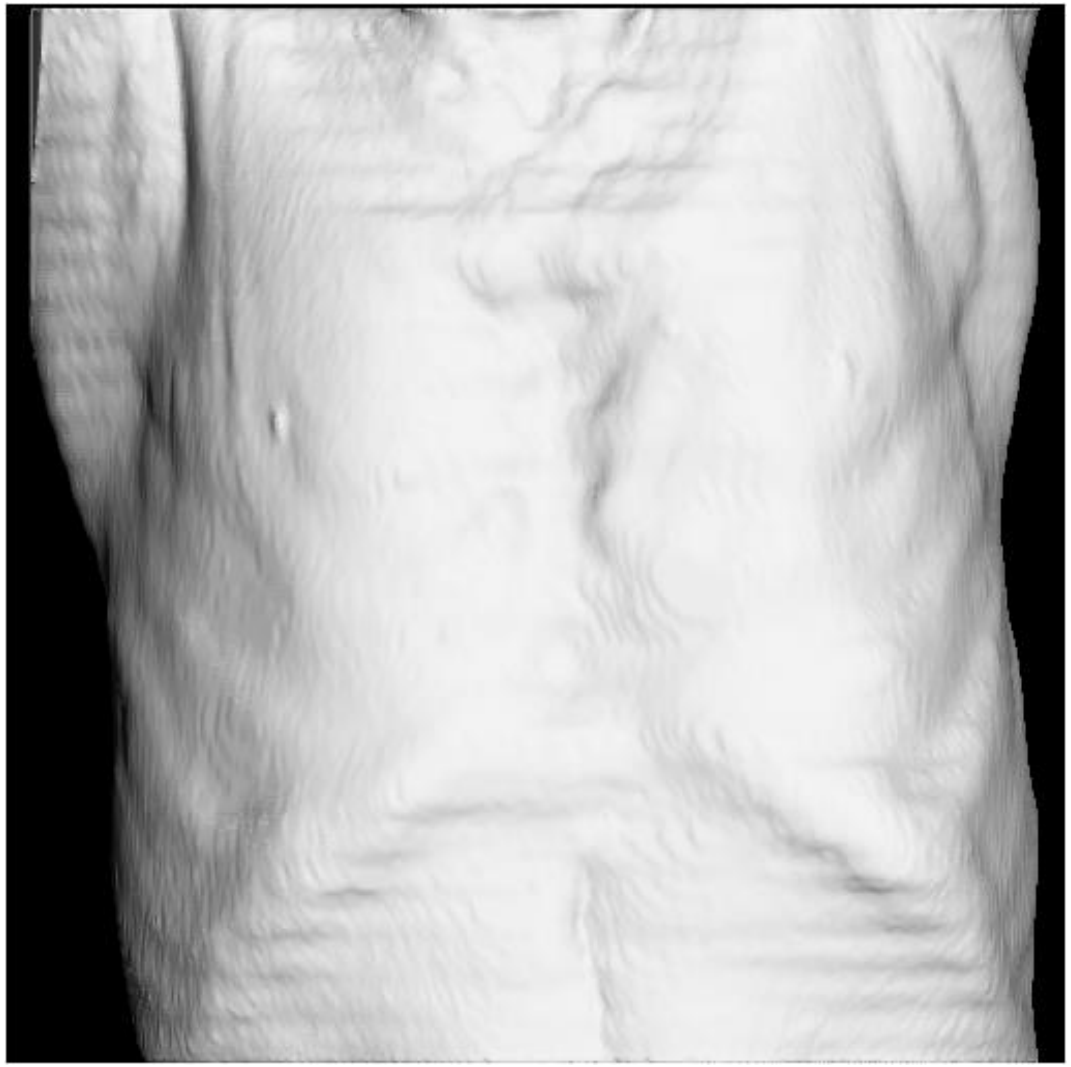
[a]Icahn School of Medicine at Mount Sinai, One Gustave Levy Place, Box 1234, New York, New York 10029, United States
[b]Early Diagnosis and Treatment Research Foundation, PO Box 1609, New York, New York 10021-0044, United States
[c]Roche Products Limited (Pharmaceuticals), Hexagon Place, 6 Falcon Way, Shire Park, Welwyn Garden City AL7 1TW, United Kingdom
[d]F.Hoffmann-La Roche, Grenzacherstrasse 124, Basel 4070, Switzerland
[e]Kitware Inc., 28 Corporate Drive, Clifton Park, New York 12065, United States
[f]Accumetra LLC, 7 Corporate Drive, Clifton Park, New York 12065, United States


(a)
(b)
(c)
(d)
(e)
(f)


50.0
40.0
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
Percent Error (%) in Volume
3056.9
2856.9
2656.9
2456.9
2256.9
2056.9
1856.9
1656.9
1456.9
1256.9
Volume
Max-Q3
Q3-Median
Median-Q1
Q1-Min
16 or less (n=12)
more than 16 (n=17)
16 or less (n=0)
more than 16 (n=6)
16 or less (n=33)
more than 16 (n=17)
Manufacturer A
Manufacturer B
Manufacturer C


A
mA: 25
kVp: 120
Thick: 1.25 mm
Kernel: LUNG

Guess which scanner?



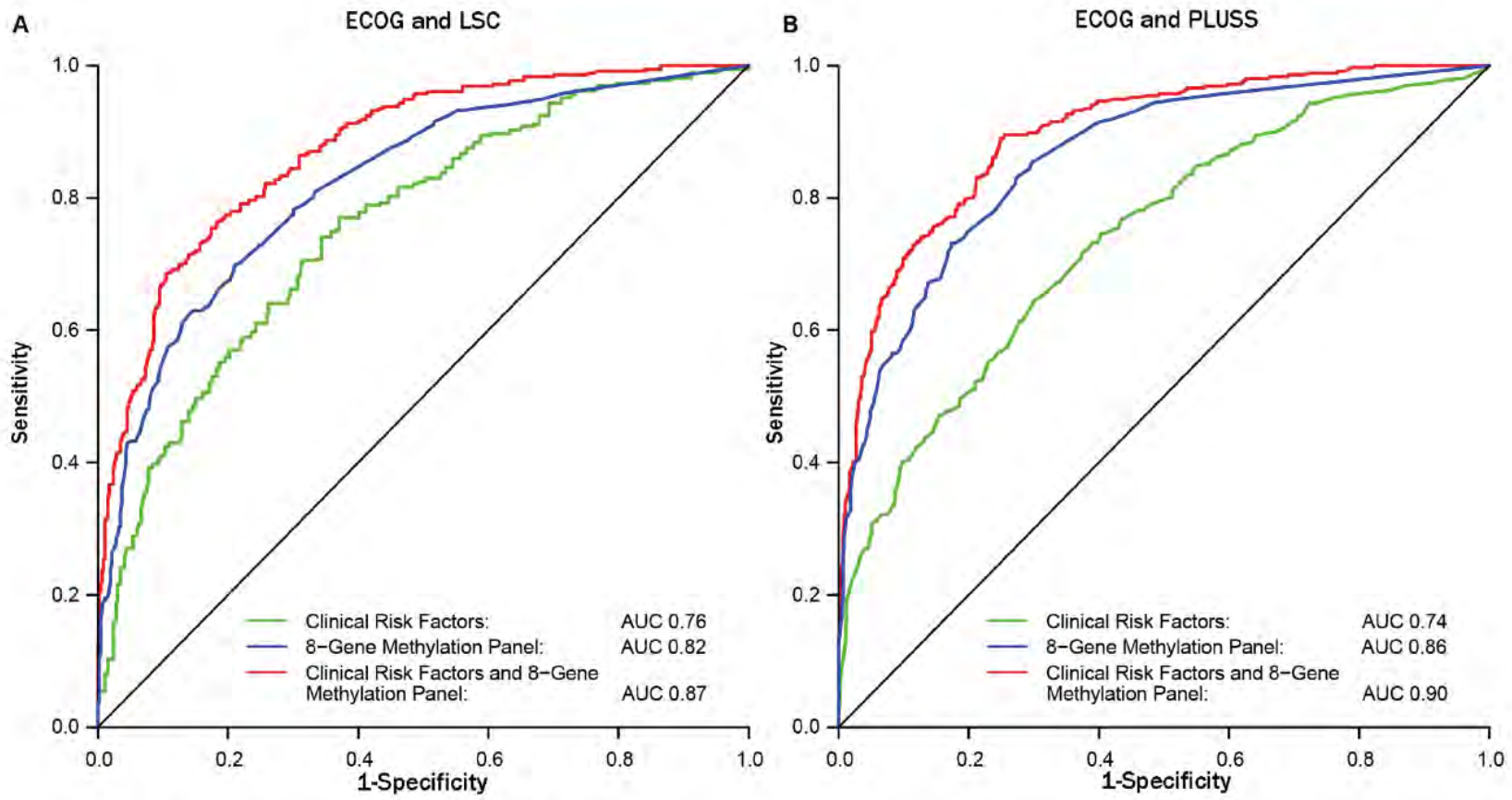


Figure 1: ROC curves for comparing the sensitivity and specificity for the eight-gene methylation panel with and without clinical risk factors between ECOG-ACRIN and LSC (A) or PLuSS (B) for classifying lung cancer risk.

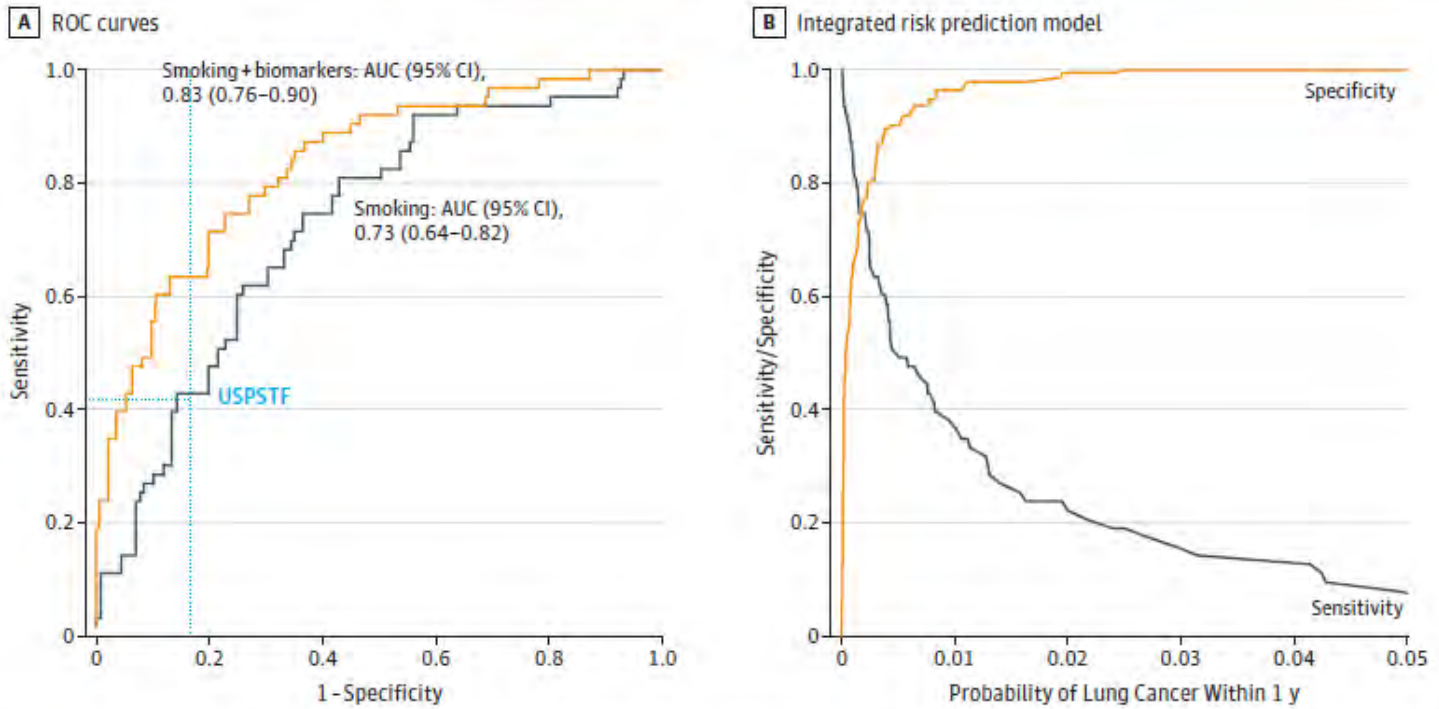
The addition of methylation biomarkers to clinical variables significantly improved lung cancer prediction accuracy.

Study Populations: ECOG-ACRIN5597 trial, the Lovelace Smokers Cohort and the PLuSS cohorts

Leng et al. Gene methylation biomarkers in sputum as a classifier for lung cancer risk. Oncotarget. 2017;8(38): p63978-63985.

Integrated risk prediction model that combined smoking exposure with the biomarker score yielded an AUC of 0.83 compared with 0.73 for a model based on smoking exposure alone (p=0.003).

Figure 3. Receiver Operating Characteristic (ROC) Curve Analysis in the Validation Study (European Prospective Investigation Into Cancer and Nutrition [EPIC] and Northern Sweden Health and Disease Study [NSHDS], Ever Smokers)



A, ROC curve analysis in the validation study (EPIC and NSHDS ever-smoker participants who received a diagnosis of lung cancer within 1 year after blood collection) for 2 risk prediction models: a model that used smoking variables only (smoking) and an integrated model with the smoking variables and the biomarker score combined (smoking + biomarkers). AUC indicates area under the curve; USPSTF, US Preventive Services Task Force. The horizontal dashed line indicates sensitivity and the vertical dashed line, specificity. B, Sensitivity and specificity in relation to the probability of lung cancer within 1 year predicted by the integrated model.

Study Populations: the European Prospective Investigation into Cancer and Nutrition (EPIC) and the Northern Sweden Health and Disease Study (NSHDS)

Integrative Analysis of Lung Cancer Etiology and Risk (INTEGRAL) Consortium for Early Detection of Lung Cancer. Assessment of Lung Cancer Risk on the Basis of a Biomarker Panel of Circulating Proteins. JAMA Oncol. 2018; 4(10):e182078.

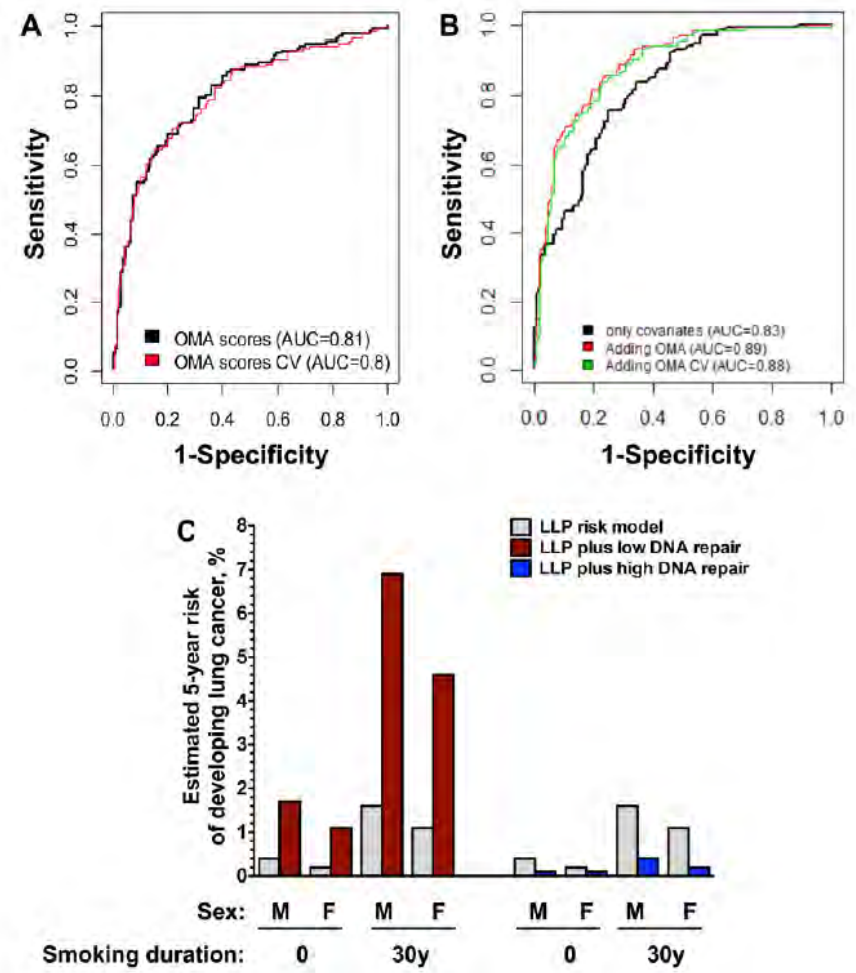


Figure 4. Receiver operating characteristic (ROC) curve of the sensitivity and specificity of the DNA repair score in lung cancer risk and examples of its added value. **A)** ROC curve for the DNA repair scores obtained in the current study (**black curve**) and after cross-validation (**red curve**). **B)** ROC curve for the covariates age, sex, and smoking status (**black curve**) and after adding the DNA repair score to these basic covariates before (**red curve**) and after (**green curve**) cross-validation. **C)** Estimated added value of the DNA repair score to lung cancer risk estimates based on the Liverpool Lung Project (LLP) Risk Model. Estimates are presented for a man and a woman, age 65 years, who are either never smokers or smoked for 30 years. **Gray columns** represent the 5-year risk according to the Lung Project Risk model. The effect of having a low DNA repair score of 5th percentile or less (**red columns**), or a high DNA repair score of 75th percentile or greater (**blue columns**) are presented. Data were taken from Supplementary Table 4 (available online). APE1 = apurinic/apyrimidinic endonuclease 1; AUC = area under the curve; CV = cross-validated; F = female; M = male; MPG = methylpurine DNA glycosylase; OGG1 = 8-oxoguanine DNA glycosylase; OMA = OGG1, MPG, and APE1.

Combining DNA Repair score with age and smoking can substantially improve lung cancer risk prediction (p=0.0002)

Study Populations: Patients from Royal Papworth Hospital (Cambridge, UK) and Cambridge BioResource.

Paz-Elizur et al. DNA repair biomarker for lung cancer risk and its correlation with airway cells gene expression. JNCI Cancer Spectrum. 2020;4(1):pkz067

median
The Imaging Phenomics Company®

Press release – For immediate release
September 06, 2021 – 05:45 pm CEST

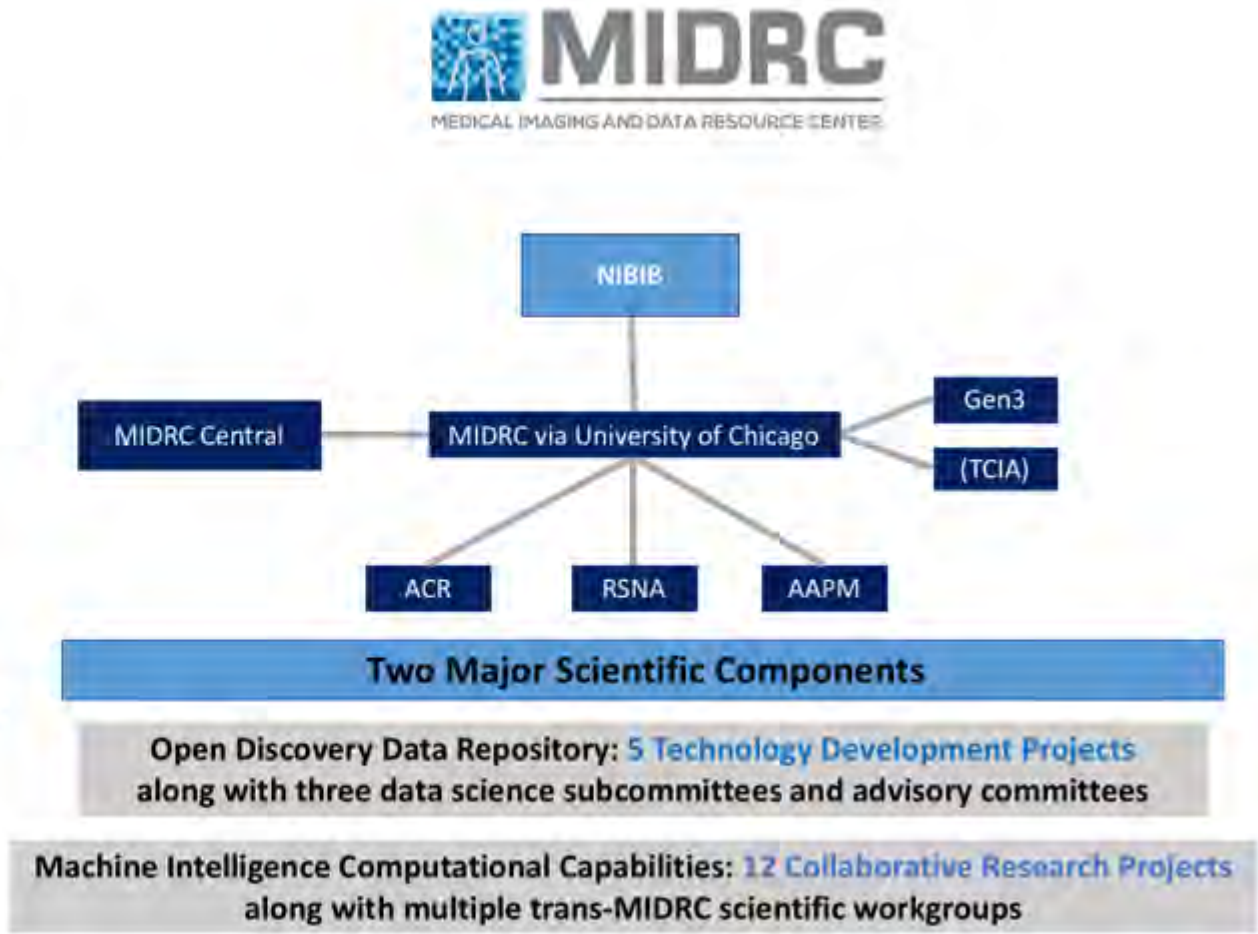
Median Technologies announces outstanding performance for its iBiopsy® Lung Cancer Screening CADx[1] to accurately characterize malignant vs benign lung nodules based on a large-scale patient cohort

- Results show cutting-edge performance of 95.2% sensitivity and 95.7% specificity for lung nodule characterization that could significantly impact lung cancer screening programs adoption.
- The large-scale study is based on a cohort of 1,696 patients with a total of 15,608 lung nodules.
- Further results on a fully automated end-to-end lung cancer screening CADe/CADx including nodule detection and characterization are expected in Q4, 2021.

The performance of iBiopsy® CADx for the characterization of lung nodules shows an AUC of 0.991 and an outstanding sensitivity of 95.2% for a specificity of 95.7%

Study Population: the National Lung Screening Trial cases (NLST)

- A multi-group two-year NIBIB-funded ($20M) project that includes:
  - AAPM, ACR, and RSNA, as well as 20 other institutions
  - Imaging and data commons through technology development projects
  - Initial research projects to expedite translation of AI from scientific findings and technical resources to public dissemination and clinical benefit



MIDRC.org

Rapid Response to COVID-19 Pandemic

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER
Established August 21, 2020

University of Chicago NIBIB Contract PI: Maryellen Giger

American Association of Physicists in Medicine (AAPM) PIs:

- Maryellen Giger (University of Chicago & AAPM Data Science Committee Chair)
- Paul Kinahan (University of Washington & AAPM Research Committee Chair)

Radiological Society of North America (RSNA) PIs:

- Curtis Langlotz (Stanford University & RSNA Board Liaison for IT & Annual Meeting)
- Adam Flanders (Thomas Jefferson University & Member RSNA CDE Committee)

American College of Radiology (ACR) PIs:

- Etta Pisano (ACR Chief Research Officer & Harvard University)
- Michael Tilkin (ACR Chief Information Officer)

Gen3 PI: Robert Grossman

NIH National Institute of Biomedical Imaging and Bioengineering

THE UNIVERSITY OF CHICAGO

AAPM AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE

ACR AMERICAN COLLEGE OF RADIOLOGY

RSNA Radiological Society of North America

MIDRC: Technology Development Projects

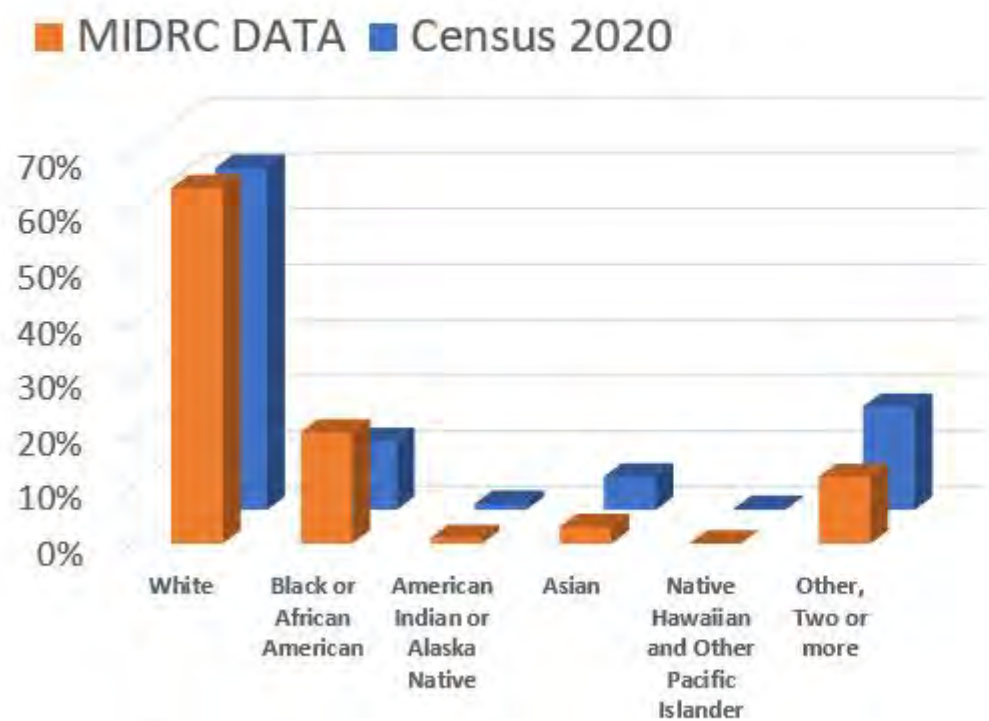
The **MIDRC infrastructure and processes** is being created through five **Technology Development Projects**, which will be conducted collaboratively:

1. Creating an open discovery platform for COVID-19 imaging and associated data (**led by RSNA**).
2. Creating a real-world testing and implementation platform with direct real-time connections to health care delivery organizations (**led by ACR**).
3. Developing and implementing quality assurance and evaluation procedures for usage across the MIDRC (**led by AAPM**).
4. Enabling data intake, access and distribution via a world-facing data commons portal (**led by all three plus Gen3**).
5. Linking the MIDRC to other clinical and research data registries (**led by all three plus Gen3**).

Three MIDRC Data Science Subcommittees

- DSIT - Data Standards and Information Technology Subcommittee
  - led by RSNA
- DPP - Data Policy and Procedures Subcommittee
  - led by ACR
- DQH - Data Quality and Harmonization Subcommittee
  - led by AAPM

Collection and curation of diverse imaging data



Diversity assessment as of May 2021

Contributions are coming from 23 states

Imaging data in the pipeline from data contribution agreements to ingestion to curation and then to release on the Gen3 output portal

Radiology

ORIGINAL RESEARCH • SPECIAL REPORT

Ethics of Using and Sharing Clinical Imaging Data for Artificial Intelligence: A Proposed Framework

David B. Larson, MD, MBA • David C. Magnus, PhD • Matthew P. Lungren, MD, MPH • Nigam H. Shah, MBBS, PhD • Curtis P. Langlotz, MD, PhD

"After clinical data are used to provide care, the primary purpose for acquiring the data is fulfilled. At that point, clinical data should be treated as a form of public good, to be used for the benefit of future patients."

https://pubs.rsna.org/doi/pdf/10.1148/radiol.2020192536

Cardiothoracic Imaging

The importance of low-dose CT screening to identify emphysema in asymptomatic participants with and without a prior diagnosis of COPD☆

David Steiger [a], M. Faisal Siddiqi [a], Rowena Yip [b], David F. Yankelevitz [b], Claudia I. Henschke [b,c,*], On behalf of theI-ELCAP investigators

[a] Division of Pulmonary, Critical Care and Sleep Medicine, Icahn School of Medicine, Mount Sinai St. Lukes, Mount Sinai West, Mount Sinai Beth Israel, New York, NY, United States of America
[b] Department of Radiology, Icahn School of Medicine, Mount Sinai Hospital, New York, NY, United States of America
[c] Phoenix Veterans Health Care System, Phoenix, AZ, United States of America

ARTICLE INFO

Keywords:
Emphysema
COPD
CT screening
Lung cancer

ABSTRACT

Purpose: Chronic Obstructive Pulmonary Disease (COPD) includes chronic bronchitis, small airways disease, and emphysema. Diagnosis of COPD requires spirometric evidence and may be normal even when small airways disease or emphysema is present. Emphysema increases the risk of exacerbations, and is associated with all-cause mortality and increased risk of lung cancer. We evaluated the prevalence of emphysema in participants with and without a prior history of COPD.

Methods: We reviewed a prospective cohort of 52,726 subjects who underwent baseline low dose CT screening for lung cancer from 2003 to 2016 in the International Early Lung Cancer Action Program.

Results: Of 52,726 participants, 23.8%(12,542) had CT evidence of emphysema. Of these 12,542 participants with emphysema, 76.5%(9595/12,542) had no prior COPD diagnosis even though 23.6% (2258/9595) had moderate or severe emphysema. Among 12,542 participants, significant predictors of no prior COPD diagnosis were: male (OR = 1.47, $p < 0.0001$), younger age ($OR_{age10} = 0.72$, $p < 0.0001$), lower pack-years of smoking ($OR_{10pack\text{-}years} = 0.90$, $p < 0.0001$), completed college or higher (OR = 1.54, $p < 0.0001$), no family history of lung cancer (OR = 1.12, $p = 0.04$), no self-reported cardiac disease (OR = 0.76, $p = 0.0003$) or hypertension (OR = 0.74, $p < 0.0001$). The severity of emphysema was significantly lower among the 9595 participants with no prior COPD diagnosis, the OR for moderate emphysema was $OR_{moderate} = 0.58(p = 0.0007)$ and for severe emphysema, it was $OR_{severe} = 0.23(p < 0.0001)$.

Conclusion: Emphysema was identified in 23.8% participants undergoing LDCT and was unsuspected in 76.5%. LDCT provides an opportunity to identify emphysema, and recommend smoking cessation.

Pathway to development of a useful database

- ‘Legacy’ databases
  - CAC (3 mm slice thickness)
  - NLST
  - Osteoporosis
- Combining information from different databases
  - Imaging, molecular, genetic, proteomic, etc
- Establishing normal values for diverse population
  - Tools for measuring the same, but normal values may be different
  - How do we determine number of participants

Pathway to development of a useful database

- Obtaining data
  - Needs of the academic institution
  - Societal needs
- Releasing data
  - Who has access?

END