Memorial Sloan Kettering Cancer Center

Design, conduct, and reporting of radiomic analyses: Let's not reinvent the wheel

Chaya S. Moskowitz, Ph.D.
Department of Epidemiology & Biostatistics

Financial disclosures: None

Outline

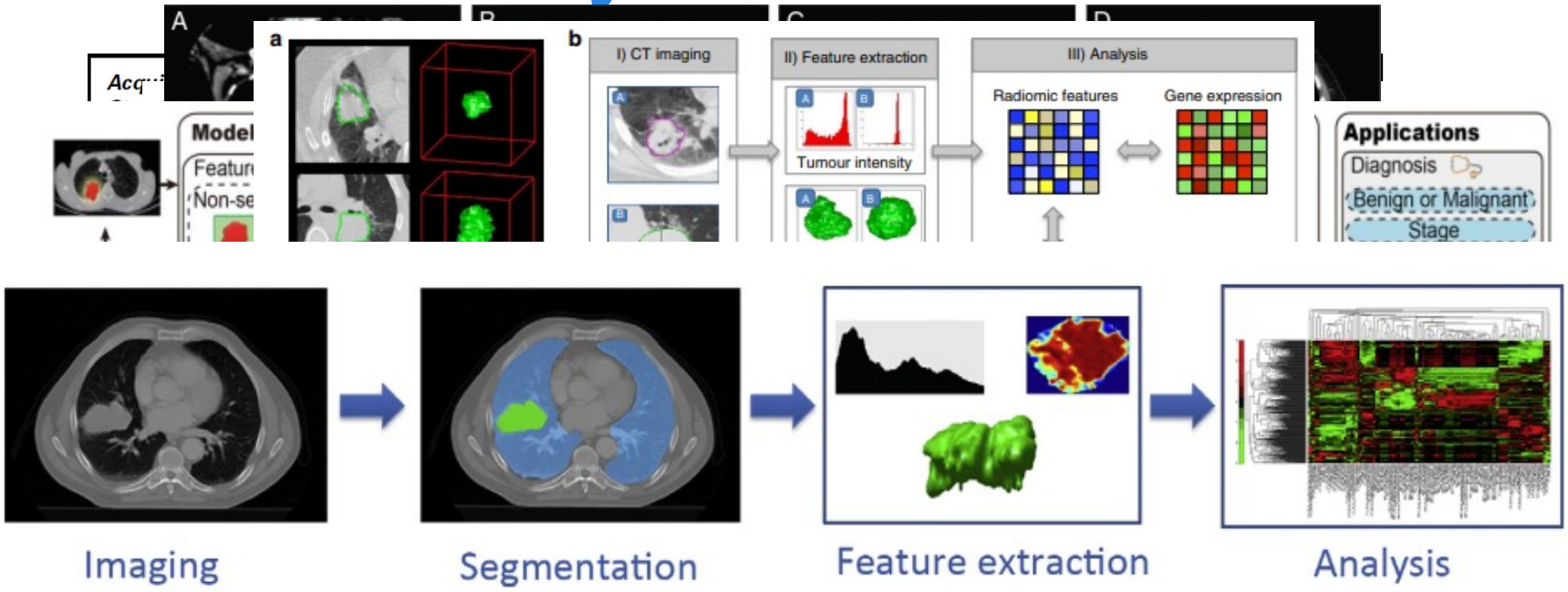
- Review of radiomics literature
- Sources of variability and bias at each step of radiomic pipeline
  - Variability and biases that are frequently pointed out
  - Additional sources and pitfalls
- How to ensure study conduct and reporting standards

Reviews of radiomics literature

- Sanduleanu and colleagues: radiomics in oncology
  - 70% of reviewed papers scored < 30% of ideal quality score
- Park and colleagues: radiomics in oncology papers published in journals with impact factors > 7.0
  - Mean quality score was 26% of ideal score
- Roberts and colleagues: radiomics in COVID-19 papers
  - 88% of papers had a high risk of bias
  - None of the models had potential for clinical use

Sanduleanu et al., Radiother Oncol 2018
Park et al., Eur Radiol 2020
Roberts et al., Nat Mach Intell 2021

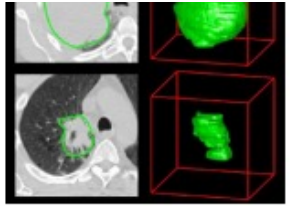
Radiomics analysis workflow



Lambin et al., EJC 2012

Liu et al.,

Trebeschi e

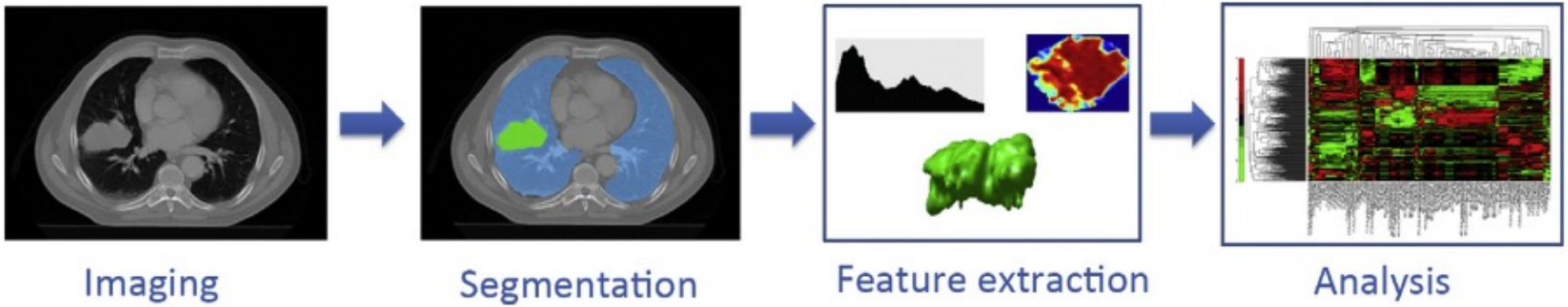


Nie et al., Int J Radiat Oncol Biol Phys 2017

Memorial Sloan Kettering Cancer Center

Radiomics analysis workflow

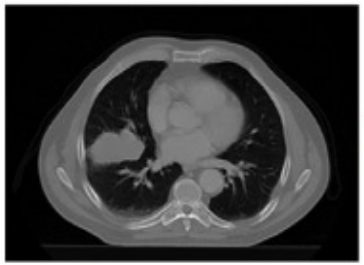
- Scanners
- Acquisition parameters
- Processing



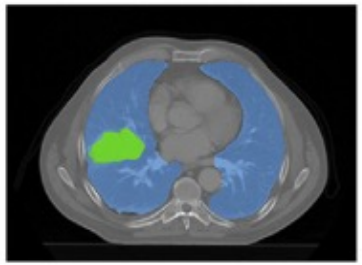
Lambin et al., EJC 2012

Radiomics analysis workflow

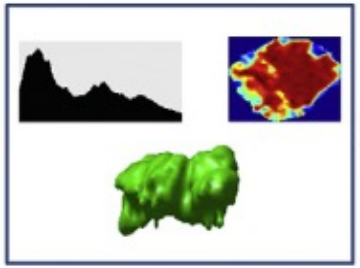
- Manual
- Semi-automatic
- Automatic



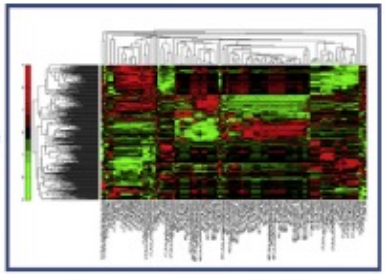
Imaging



Segmentation



Feature extraction

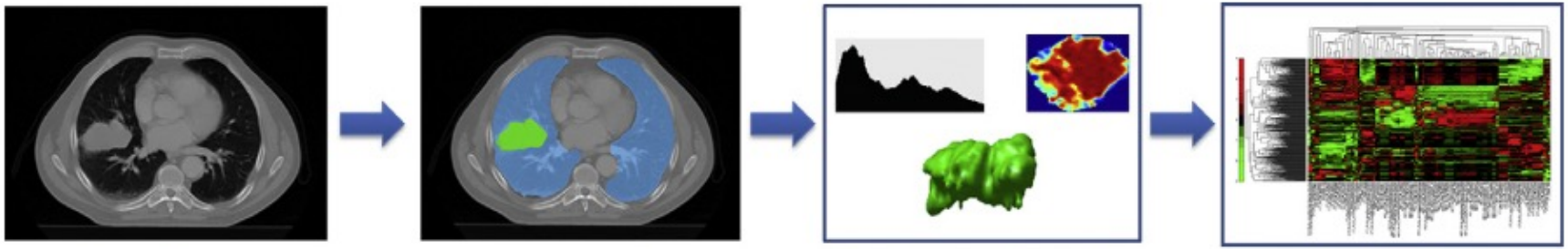


Analysis

Lambin et al., EJC 2012

Radiomics analysis workflow

- Computation methods
- Nomenclature



Imaging → Segmentation → Feature extraction → Analysis

Lambin et al., EJC 2012

Standardization Inititatives

Quantitative Imaging Biomarkers Alliance

RSNA

NIH NATIONAL CANCER INSTITUTE
DCTD Division of Cancer Treatment & Diagnosis

CIP Cancer Imaging Program

Home | About CIP | Research Funding | Programs & Resources | Clinical Trials | Informatics

The next QIN Annual Meeting will be held January 22nd 2021. Please register here.

About QIN

Background, Mission, and Goal

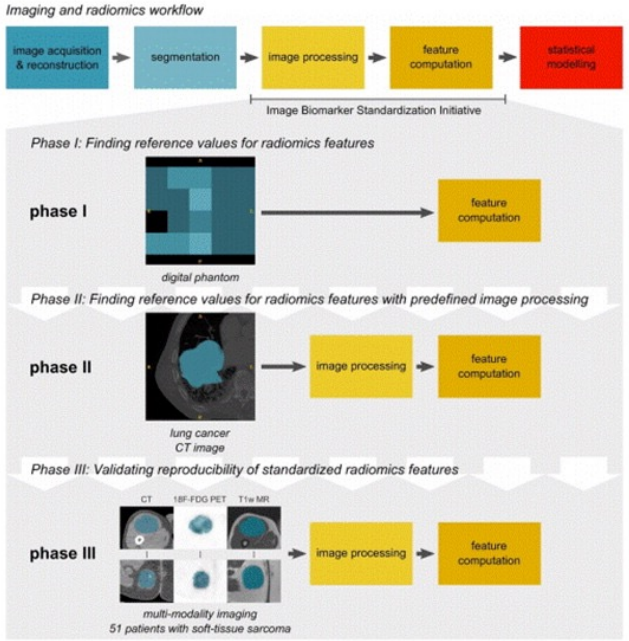
Teams

Programs & Resources | Quantitative Imaging Network (QIN)

Quantitative Imaging Network (QIN)

About the Quantitative Imaging Network (QIN)

The Image Biomarker Standardization Initiative

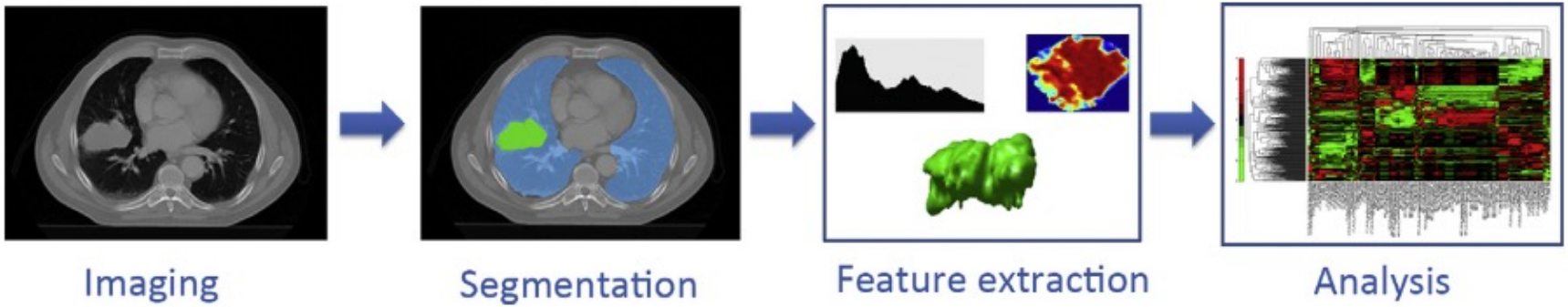


Clarke et al., Transl Oncol 2014
Sullivan et al., Radiology 2015
Zwanenburg et al., Radiology 2020

Memorial Sloan Kettering Cancer Center

Radiomics analysis workflow

- Feature selection methods
- Classification and model building methods
- Validation methods (or lack thereof)



Lambin et al., EJC 2012

Frequent Sources of Bias in Radiomic Analyses

- Bias due to overfitting
- Optimistic performance bias
- Multicollinearity
- Multiple testing

Radiomics Quality Score

Image protocol quality	Multiple segmentation	Phantom study	Imaging at multiple time points
+1 or +2	+1	+1	+1
RQS checkpoint 1 total: 2			RQS checkpoint 2 total: 3

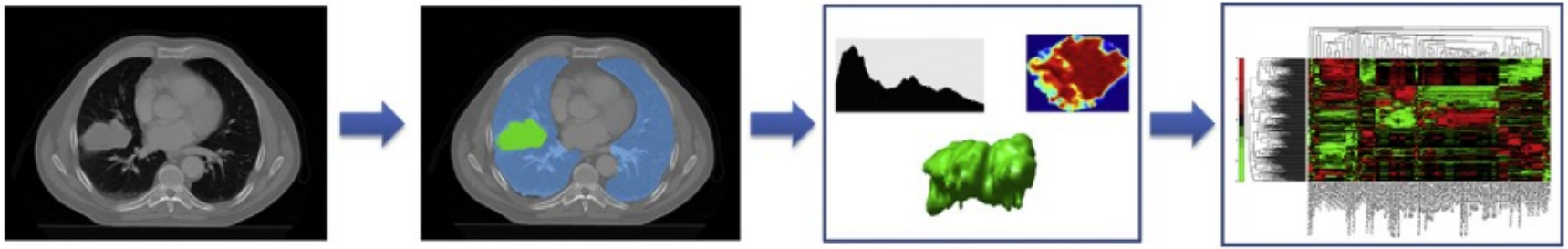
Feature reduction or adjustment for multiple testing	Multivariable analysis	Biological correlates	Cut-off analysis	Discrimination statistics	Calibration statistics	Prospective study	Validation	Comparison to 'gold standard'	Potential clinical applications	Cost-effectiveness analysis	Open science and data
-3 or +3	+1	+1	+1	+1 or +2	+1 or +2	+7	-5 to +5	+2	+2	+1	+1 to +4
											RQS checkpoint 3 total: 31

RQS Total	36

Lambin et al., Nat Rev Clin Oncol 2017

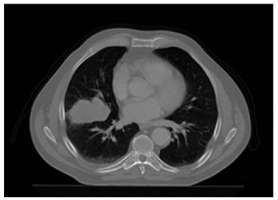
Memorial Sloan Kettering Cancer Center

Radiomics analysis workflow

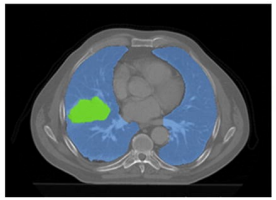


Lambin et al., EJC 2012

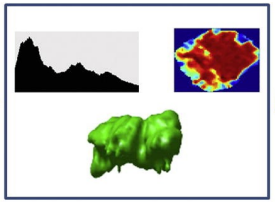
Radiomics analysis workflow



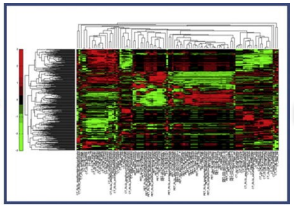
Imaging



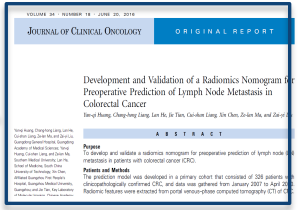
Segmentation



Feature extraction



Analysis



VOLUME 34 · NUMBER 18 · JUNE 20, 2016

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer

Tian-yi Huang, Chang-long Liang, Lan He, Jie Tian, Cui-shan Liang, Xin Chen, Zi-lan Ma, and Zai-yi

ABSTRACT

Purpose

To develop and validate a radiomics nomogram for preoperative prediction of lymph node metastasis in patients with colorectal cancer (CRC).

Patients and Methods

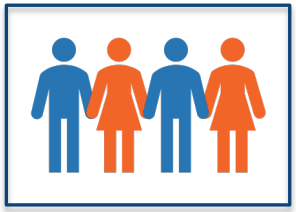
The prediction model was developed in a primary cohort that consisted of 326 patients with clinicopathologically confirmed CRC, and data was gathered from January 2007 to April 20

Radiomic features were extracted from portal venous-phase computed tomography (CT) of

Reporting & interpretation

Prospective study

Retrospective study



Study design

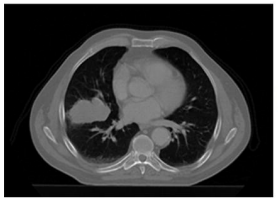
Memorial Sloan Kettering Cancer Center

Study design

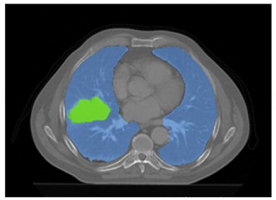
Elements of study design include:

- Research hypotheses
- Study objectives
- Endpoints
- Reference information
- Patient population
- Inclusion/exclusion criteria for images
- Reader study design if applicable

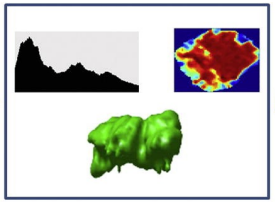
Potential Sources of Variability and Bias



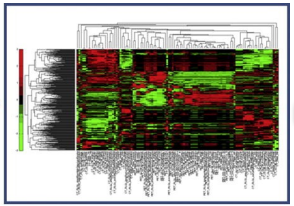
Imaging



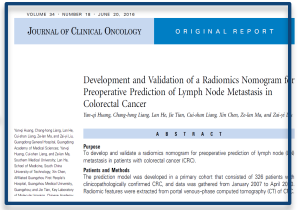
Segmentation



Feature extraction



Analysis



JOURNAL OF CLINICAL ONCOLOGY ORIGINAL REPORT

Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer

ABSTRACT

Purpose

Patients and Methods

Reporting & interpretation

Prospective study

Retrospective study

Study design

- Outcome / reference standard
- Patient / Image selection
- Confounding

Memorial Sloan Kettering Cancer Center

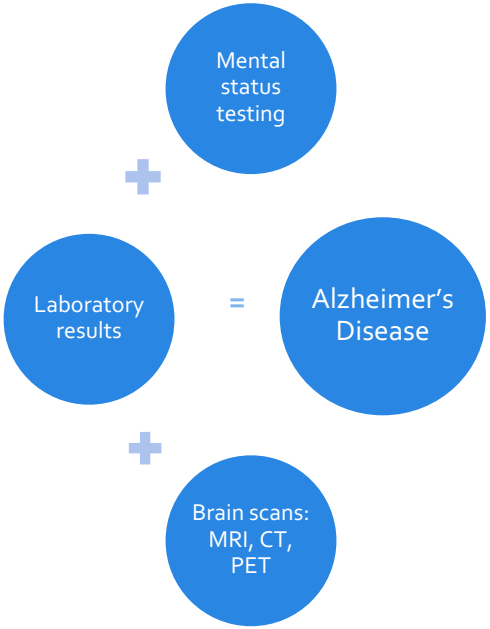
Outcome

- In diagnostic accuracy studies, outcome = reference or gold standard
- In radiomic analysis, there might not be a reference standard for a feature or signature; the outcome could be presence of abnormal condition either at the time the image is acquired or in the future (e.g. overall or progression-free survival)
- Essential to carefully define outcome
- Definition of the outcome should not rely on information from the imaging modality producing the radiomic features
  - If it does → incorporation bias

Incorporation Bias in Diagnostic Testing Framework

- The outcome uses information from the images being analyzed
- Occurs most often when clinical judgment is needed to determine the outcome
- Leads to overestimating measures of accuracy

Diagnosing Alzheimer's Disease



Ransohoff and Feinstein, NEJM 1978
Whiting et al., Ann Intern Med 2004

Memorial Sloan Kettering Cancer Center

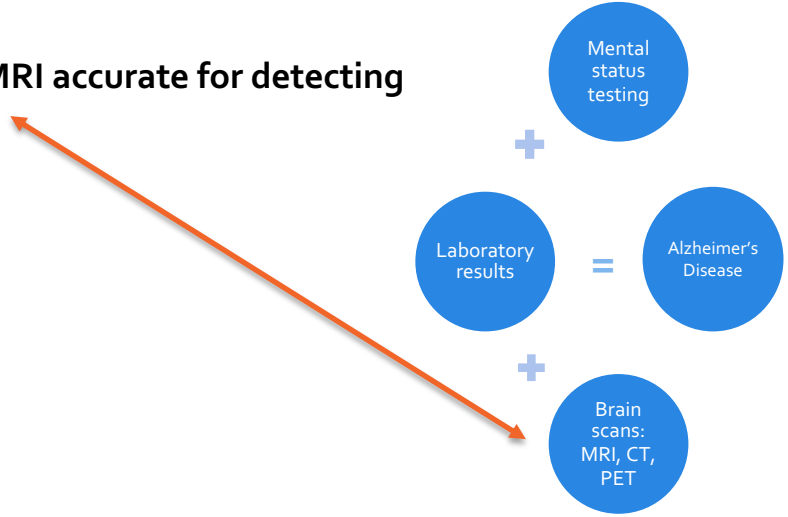
Incorporation Bias in Diagnostic Testing Framework

- The outcome uses information from the images being analyzed
- Occurs most often when clinical judgment is needed to determine the outcome
- Leads to overestimating measures of accuracy

Is MRI accurate for detecting Alzheimer's Disease?

EQUIVALENT TO

Is MRI accurate for detecting **?**



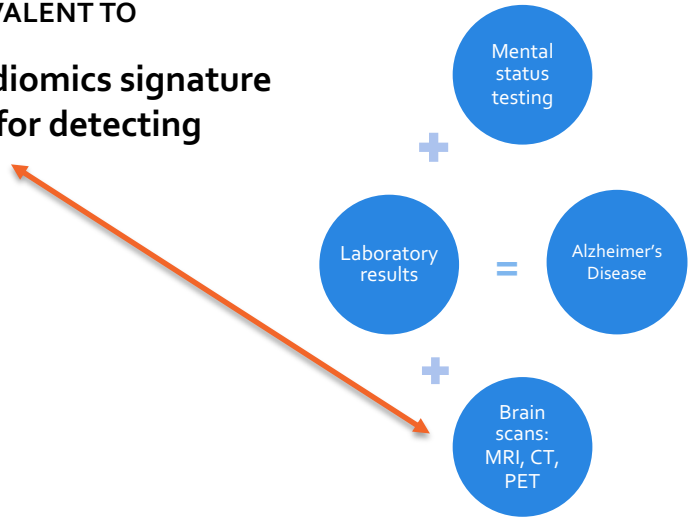
Incorporation Bias in Radiomics

- Bias is greatest when one or more of the features are vital for identifying the condition
- Bias may still be present if the features under study play less of a role in identifying the condition
- Difficult to understand degree of overestimation

Is MRI radiomics signature accurate for detecting Alzheimer's Disease?

EQUIVALENT TO

Is MRI radiomics signature accurate for detecting



?

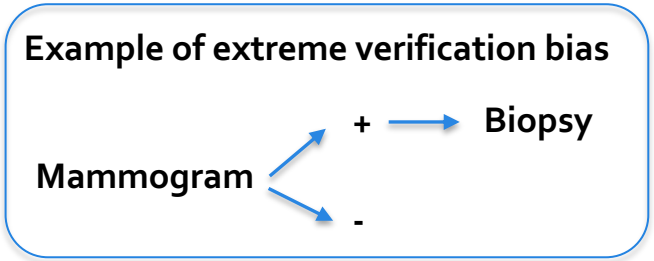
Who to include: Spectrum Bias

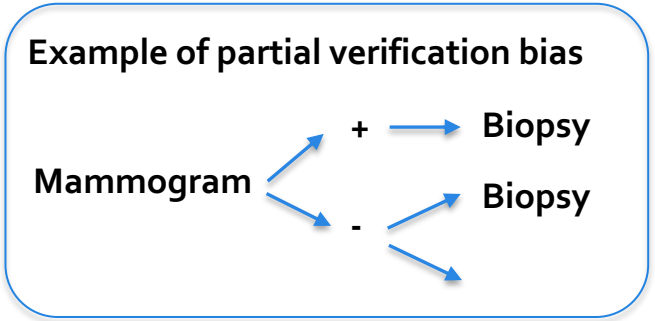
- Study data not fully representative of the population of interest
  - Patients with outcome have severe disease or health conditions that are more obvious
  - “Healthy” patients are more healthy than typical patients
- Leads to overestimating accuracy
- Example
  - Using scans from patients with osteoarthritis referred to specialty care to develop a model for use in a primary care population
- Model is affected is by differences in patient characteristics or settings (spectrum effect)

Ransohoff and Feinstein, NEJM 1978
Williams, Int Musculoskelet Med 2010

Verification Bias in Diagnostic Testing Framework

- Results from non-random selection of images
- Study is limited to images with the outcome ascertained
- Extreme verification bias: *only* images that are suspicious for disease have the outcome ascertained and are included in the study
- Partial verification bias: a non-random portion of images have the outcome ascertained
- Direction of bias is difficult to know
  - Often results in increased sensitivity
  - Can result in increased or decreased specificity
  - AUC?



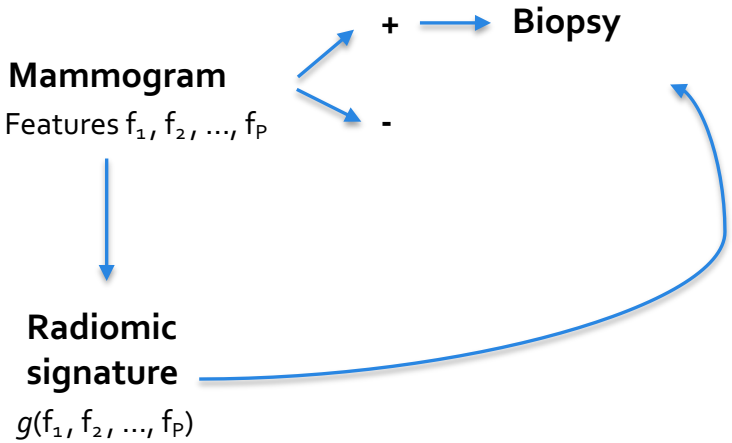


Begg and McNeil, Radiology 1988
Kohn et al., Acad Emerg Med 2013

Memorial Sloan Kettering Cancer Center

Verification Bias in Radiomics

- Path may be more indirect than in classical sense
- Bias may be diluted, but it is still present
- Bias is likely to be greatest when one or more of the imaging features are vital for informing decision to verify outcome
- Very difficult to understand how it affects our evaluation of model performance



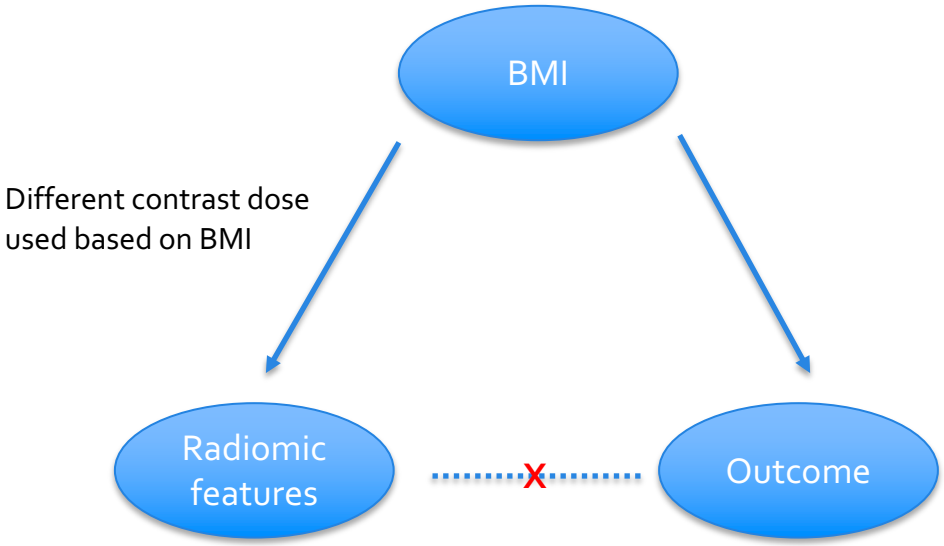
Bias Due To Confounding

- “Mixing of effects”
- Confounding is a distortion of the estimated association between radiomic features and the outcome that occurs because the radiomic features are entangled with another factor associated with the outcome.
- Can lead to missing an association that is present or falsely finding an association when none is present
- Important to collect information on confounders and adjust for it in the analysis

Weiss, Clinical Epidemiology 2006

Source of Confounding in Radiomics Studies

- Different imaging protocols based on clinical factors
- Imaging differences (protocol variations or artifacts) affect feature measurement
- If they are related to clinical factors associated with the outcome, this can lead to confounding



Moving Forward

What do we need to improve the conduct and reporting of radiomic studies?

I haven’t said anything new here!

Conclusions

- Potential of radiomic studies weighed down by bias and variability
- Many sources of bias and variability have been identified in radiomics and other literature
- Involving experts from necessary disciplines can help
- Do we need guidelines for conduct and reporting specifically for radiomic studies?

Thank You!

Acknowledgments

Amber Simpson	Queen's University
Mattea Welch	Princess Margaret Cancer Centre
Brenda Kurland	ERT

Follow-up questions or thoughts?
e-mail: moskowc1@mskcc.org
Twitter: @ChayaSMoskowitz

Memorial Sloan Kettering Cancer Center