The Oncoshare Project

Allison W. Kurian, MD, MSc
Associate Professor of Medicine, Epidemiology & Population Health
Director, Women's Clinical Cancer Genetics Program
Stanford University School of Medicine

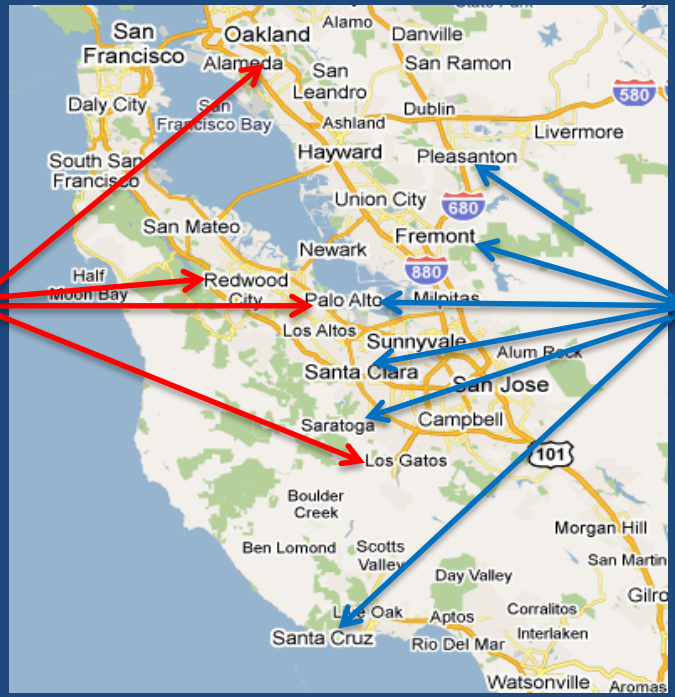
Background and Goals

- Cancer data sources are fragmented
  - Registries: demographics, survival; limited diagnostics, treatment
  - Electronic medical records: clinical details, but unstructured text
  - Send-out laboratory test results are not integrated into EMRs
- Most data sources do not span healthcare settings
- We aimed to integrate data sources across healthcare systems
- Goals: to enable quality assessment and inform interventions

Contributing Healthcare Systems

STANFORD HOSPITAL & CLINICS
Stanford University Medical Center

Academic, Tertiary Care
Research-Centered
Knowledge Discovery

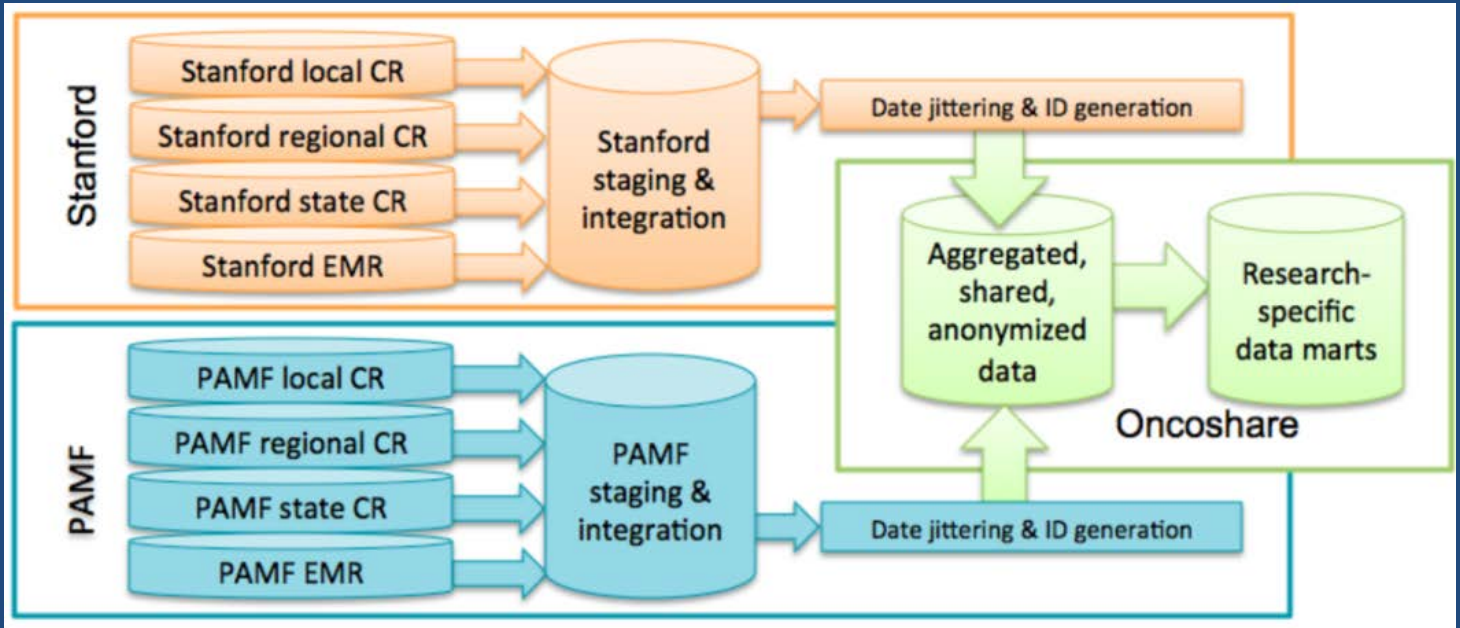


Palo Alto Medical Foundation
A Sutter Health Affiliate

Community-Based
Patient-Centered

- Two healthcare systems serving the San Francisco Bay Area

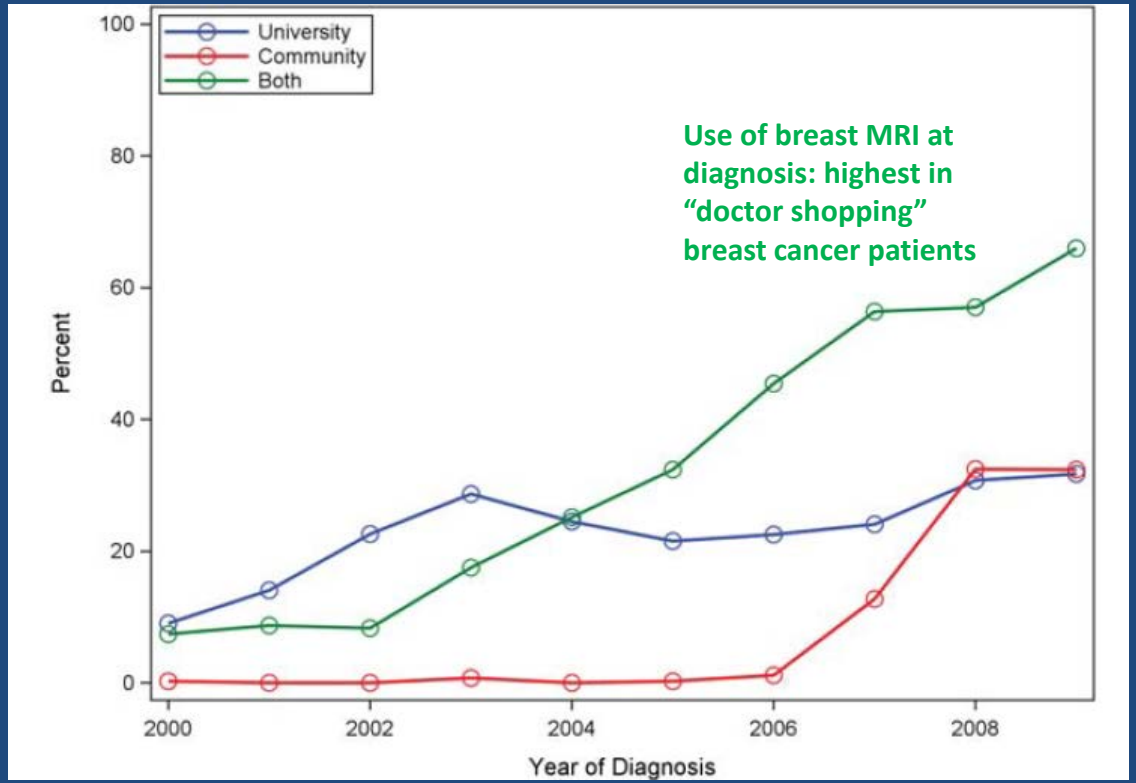
Design and Methods



Patient-level data from:
Statewide California Cancer Registry (SEER Program), EMRs from two healthcare systems and outside genomics labs

Progress to Date (July 2020): >28,000 breast cancer patients, diagnosed 2000-18
40 publications and presentations to date; website: http://med.stanford.edu/oncoshare.html

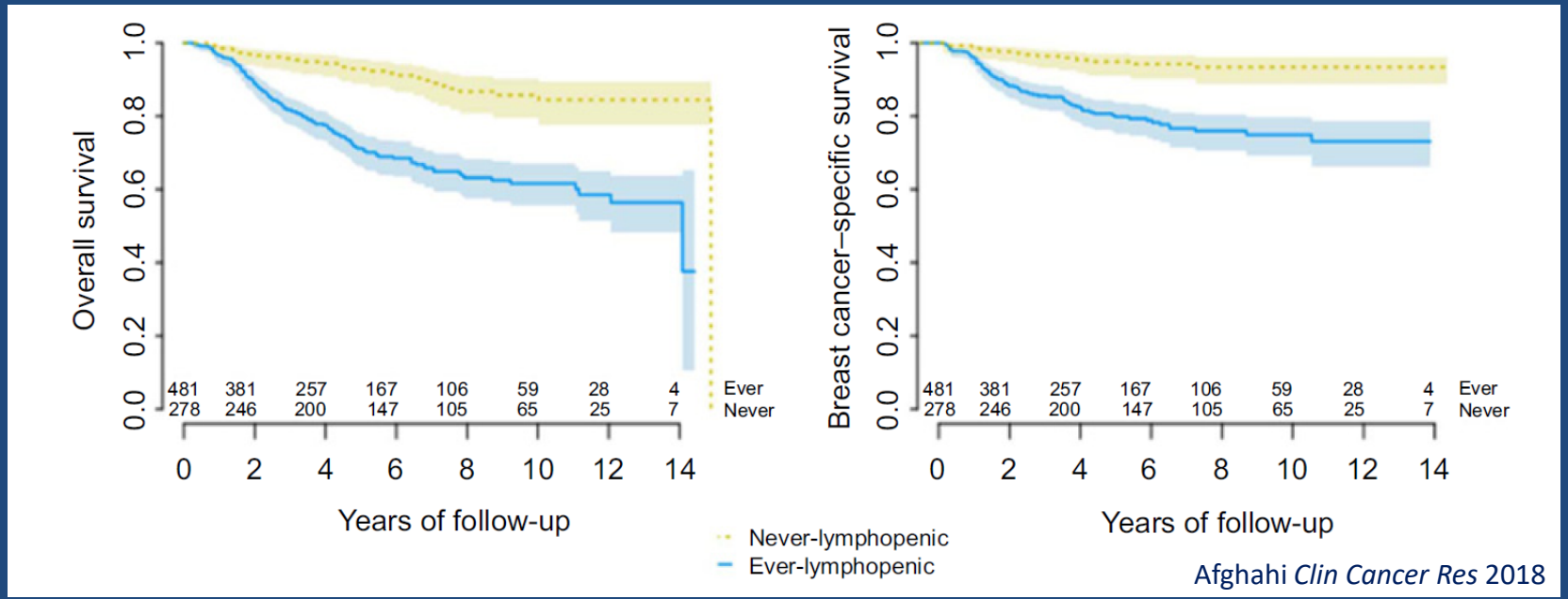
Results Example: Unwarranted Variation in Care



Kurian *Cancer* 2014; Afghahi *JCO Oncology Practice* 2016

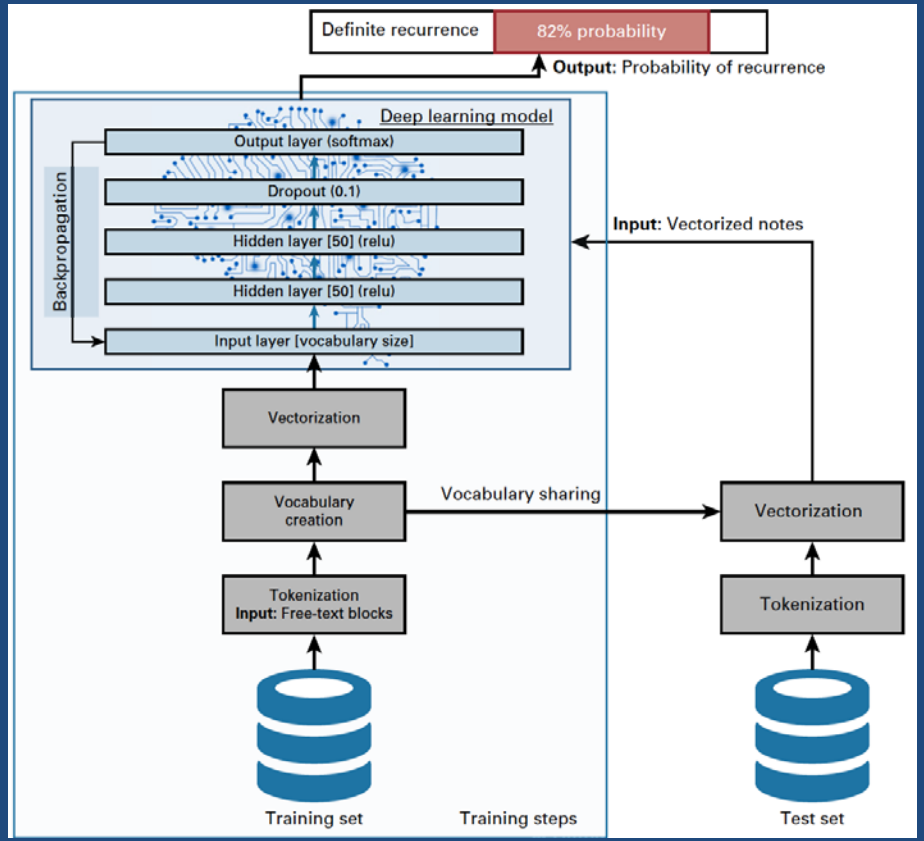
- 15% of patients seek care at both institutions in same area
- "Both" patients: no difference in prognostic factors vs. others
- "Both": more MRI, PET, bilateral mastectomy, chemotherapy, radiotherapy ($p<0.001$), but no survival difference ($p=0.66$)
- A hot spot of unwarranted variation in care; may inform targeted interventions

Results Example: Immune Function and Survival



- Triple-negative breast cancer cohort in Oncoshare: n>1,500, 1/4 died within five years of diagnosis
- Investigated peripheral absolute lymphocyte count (ALC), a measure of immune function
- Significantly worse survival if ever lymphopenic; could ALC enhance treatment effect?

Results Example: Identifying Metastatic Recurrence



Banerjee *JCO Clinical Informatics* 2019

- Distant recurrence: a clinically important endpoint not tracked by registries
- Manually annotated 894 randomly selected patient records: recurrence yes/no, date
- Used natural language processing to develop rule-based and neural network models
- Neural network model: 83% sensitivity, 73% specificity for detecting recurrence and date
- Adaptable for cancer types other than breast

Future Plans

- Integration of novel data sources
  - Tumor sequencing data from outside laboratories
  - Imaging data from EMR
  - Patient-reported data from EMR portals
- Partnering with other healthcare systems to validate approaches
  - Collaborations with Emory, MIT, Vanderbilt in U.S.
  - Koo Foundation Sun Yat Sen Cancer Center in Taiwan
- Consulting with Surveillance Epidemiology and End Results program
  - Sharing approaches that may inform SEER registry data integration

Acknowledgments

Stanford University

Douglas Blayney
Alison Callahan
Jennifer Caswell-Jin
Manisha Desai
Summer Han
Haruka Itakura
Yasmin Karimi
Albee Ling
Mina Liu
Natasha Purington
Daniel Rubin
Nigam Shah
Lidia Schapira
George Sledge
Suzanne Tamang
Susan Weber

Palo Alto Medical Foundation

Su-Ying Liang
Monique de Bruin
Dominick Frosch
Pragati Kenkare
Harold Luft

UCSF

Scarlett Gomez

Emory University

Imon Banerjee

Research Funding

BRCA Foundation
Breast Cancer Research Foundation
California Breast Cancer Research Program
Jan Weimer Faculty Chair in Breast Oncology
Richard and Susan Levy Gift Fund
Stanford Cancer Institute Developmental Research Awards
Suzanne Pride Bryan Breast Cancer Research Fund