Translating from diagnostic performance to outcomes in multi-cancer early detection testing: an evidence generation proposal

Ruth Etzioni PhD

Fred Hutchinson Cancer Center

Rosalie and Harold Rea Brown Endowed Chair

ICSN 2023

Multi-cancer blood test shows real promise in NHS study BBC

Mercy among first to offer $949 blood test that can screen for more than 50 types of cancer

HEALTH · PUBLIC HEALTH

Could a simple blood test detect cancer at an early—and more treatable—stage? The technology exists—and FDA approval may not be far off

Prevent Cancer Foundation champions introduction of Nancy Gardner Sewell Medicare Multi-Cancer Early Detection Screening Coverage Act in the House

Why so much interest?

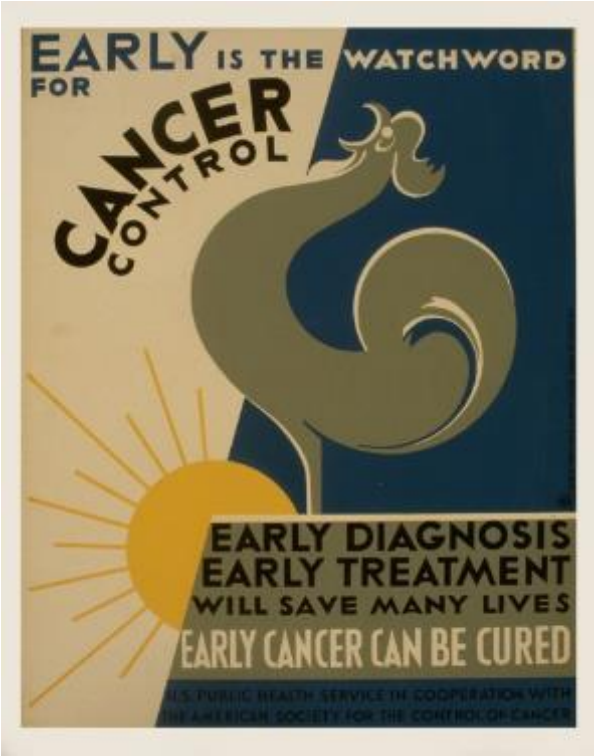
Published measures of diagnostic performance are promising*

Major technology advance

Enduring faith in the early detection solution

Marketing

1960s



2020s

Tests differ in their features, cancer targets and outputs

Features of circulating tumor DNA

- Methylation profiles
- Fragment size distributions
- Specific mutations
- Other features

Cancer targets

- Single cancers
- Specific subsets of cancers e.g. smoking related
- Large numbers (>50) cancers

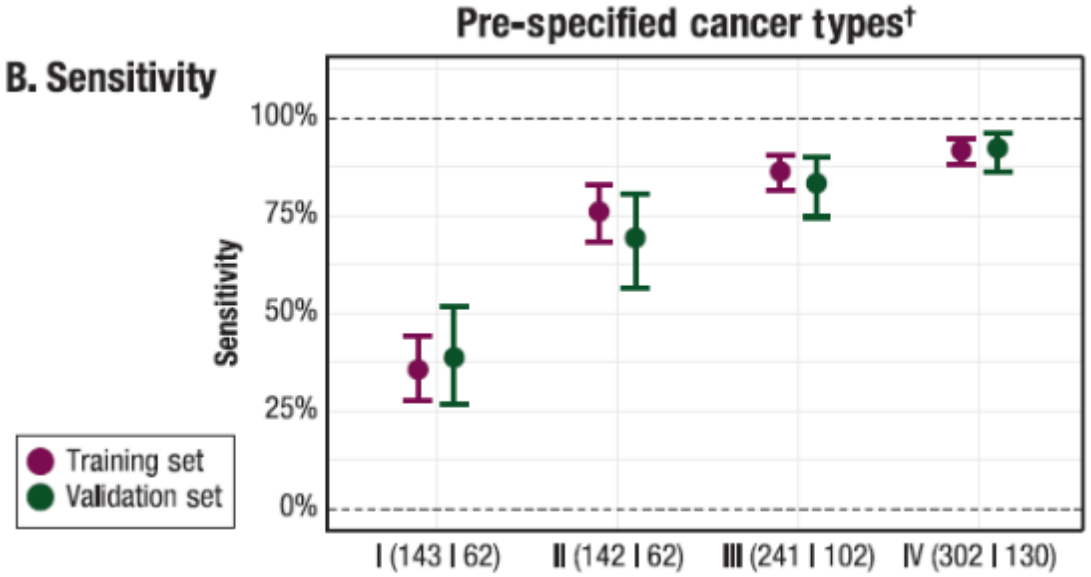
Outputs

- Predictive algorithms yield a score that is thresholded to produce result
- Most tests provide a tissue of origin but do not specify recommended workup
- One test does not provide a TOO and recommends whole-body PET/CT

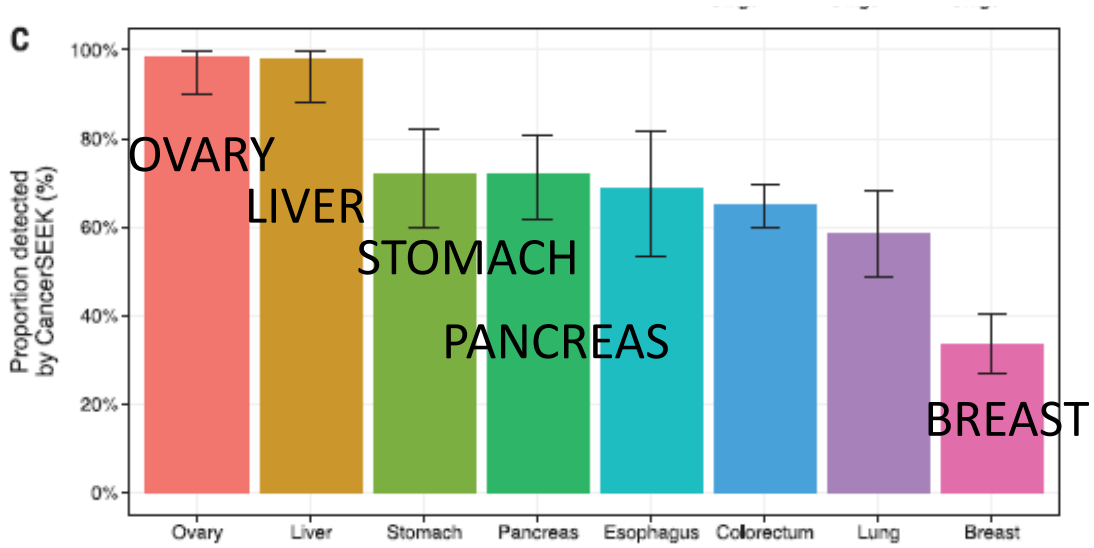
Published data about test performance

1. Tests have very high specificity - by construction
   - Algorithmic threshold selected to be conservative
2. They can find cancer when we know it is there
   - Modest performance for early-stage cases
   - Better sensitivity for late-stage cases
   - Low sensitivity for pre-cancers
3. They have high positive predictive value in a prospective setting
   - Likely due to higher prevalence of multi-cancer plus high specificity
4. They have modest sensitivity in this setting
   - To an extent lower sensitivity under prospective screening is to be expected

Sensitivity among known cases



Sensitivity by stage (overall 67.3%)



Sensitivity by cancer (median 70%)

Liu et al Annals of Oncology 2020

Cohen *et al.*, *Science* **359**, 926–930 (2018)

Positive predictive value in screened cohort

Beer et al ASCO 2021

MCED TEST DETECTED BROAD RANGE OF CANCER SIGNALS, INFORMING DIAGNOSTIC WORKUP, WITH 45% PPV

Diagnostic Workup

Table 1. Diagnostic Workup Procedures Per Participant with Diagnostic Resolution

	Median (Q1, Q3)		
	True Positive (n = 27[a])	False Positive (n = 36)	Total (n = 63)[a]
All Imaging Tests/Invasive Procedures	2.0 (1.5, 3.0)	1.5 (1.0, 2.2)	2.0 (1.0, 3.0)
All Imaging Tests	1.0 (1.0, 1.5)	1.0 (1.0, 2.0)	1.0 (1.0, 2.0)
Functional[b]	1.0 (0, 1.0)	1.0 (0, 1.0)	1.0 (0, 1.0)
Anatomic[c]	1.0 (0, 1.0)	1.0 (0, 1.0)	1.0 (0, 1.0)
All Invasive Procedures	1.0 (1.0, 1.0)	0 (0, 0.2)	0 (0, 1.0)
Minimally Invasive[d]	1.0 (0.5, 1.0)	0	0 (0, 1.0)
Surgical[e]	0	0	0
Clinical Lab Tests	3.0 (1.0, 5.5)	3.0 (1.0, 6.0)	3.0 (1.0, 6.0)
Days to Diagnostic Resolution	50.0 (27.0, 76.5)	49.0 (30.2, 153.8)	50.0 (28.0, 91.0)

Abbreviations: CT, computerized tomography; MR, magnetic resonance imaging; PET, positron emission tomography; Q1, first quartile; Q3, third quartile.
[a] 2 participants with 'signal detected' MCED test result (true positives) were excluded from the diagnostic workup analysis because diagnostic testing was initiated before MCED test results were returned.
[b] Functional imaging includes PET-CT, PET-MRI, bone scan.
[c] Anatomic imaging includes CT, MRI, ultrasound, mammography, plain film X-ray (including skeletal survey).
[d] Minimally invasive procedures include esophagogastroduodenoscopy, colonoscopy, endoscopic ultrasound, endoscopic retrograde cholangiopancreatography, bronchoscopy, cystoscopy, hysteroscopy, fine needle aspiration of the thyroid gland, liver biopsy, thoracentesis, pulmonary arterio-venous malformation embolization.
[e] 4 surgical procedures were performed, 3 in true positive, 1 in false positive set.

- Most participants with diagnostic resolution (57/63, 90.5%) had at least 1 imaging test
  - Median number of imaging tests per participant was the same in true and false positive groups (**Table 1**)
- Most invasive procedures were minimally invasive (28/32 procedures, 87.5%)
  - 26/30 (86.6%) participants had only minimally invasive procedures
- There were 4 reports of study-related adverse events; 3 of anxiety and 1 bruise at venipuncture site, all were of mild severity

Test Performance

Table 2. MCED Test Performance

	≥50 y With Additional Risk	≥50 y Without Additional Risk	Total
Cancer Signal Detection, No.	n=3695	n=2934	N=6629
Detected, No. (%)	56 (1.5)	36 (1.2)	92 (1.4)
True Positive	20 (0.5)	9 (0.3)	29 (0.4)
False Positive	15 (0.4)	21 (0.7)	36 (0.5)
No Current Diagnostic Resolution	21 (0.6)	6 (0.2)	27 (0.4)
Not Detected	3639 (98.5)	2898 (98.8)	6537 (98.6)
PPV for Cancer Signal Detection, No.	n=35	n=30	n=65
% (95% CI)	57.1 (40.9-72.0)	30.0 (16.7-47.9)	44.6 (33.2-56.7)
CSO Prediction Accuracy, No.	n=19[a]	n=8[a]	n=27[a]
First CSO,[b] % (95% CI)	84.2 (62.4-94.5)	87.5 (52.9-99.4)	85.2 (67.5-94.1)
First or Second CSO,[c] % (95% CI)	100 (83.2-100)	87.5 (52.9-99.4)	96.3 (81.7-99.8)

Abbreviations: CI, confidence interval; CSO, cancer signal origin; PPV, positive predictive value.
[a] Excludes 1 participant with unknown cancer type and 1 with indeterminate CSO from the true positive set.
[b] Proportion of correctly predicted first CSO among true positive participants with determinate CSO.
[c] Proportion of correctly predicted first or second CSO among true positive participants with determinate CSO.

- Of the 6629 analyzable participants, the MCED test detected cancer signal in 92 (1.4%); 1.5% of participants with additional risk and 1.2% without (**Table 2**)
- The PPV of the MCED test for participants with cancer signal detected who achieved diagnostic resolution was 44.6% (**Table 2**)
  - PPV was 57.1% in the "additional risk" vs 30.0% in the "without additional risk" cohort

Table 3. Cancer Stage at Diagnosis Following a Positive MCED Result (n=28[a] True Positives)

Cancer Type Diagnosed	Clinical AJCC Stage[b] of New Cancers					Extent of Recurrent Cancers		First Predicted Cancer Signal Origin
	I	II	III	IV	Other	Local	Distant	
Colon or rectum				1	1 Unknown[c]			Upper GI Tract (stage IV pt) Colon/Rectum (unk pt)
Head and Neck		1		1				Head and Neck
Liver, bile duct	1		1					Liver, bile-duct
Lung			1					Lung
Lymphoid leukemia					2 NA[d]			Lymphoid Neoplasm
Lymphoma	2	3	1	2				Lymphoid Neoplasm
Ovary, peritoneum or fallopian tube			1					Uterus (ovary second CSO)
Pancreas		1						Pancreas/Gallbladder
Plasma cell neoplasm					1 NA[d]			Plasma Cell Neoplasm
Prostate				1				Indeterminate
Small intestine	1							Colon/Rectum (upper GI second CSO)
Waldenstrom macroglobulinemia					1 NA[d]			Lymphoid Neoplasm
Breast							4	3 cases Breast 1 case Breast (first CSO), lymphoid (second)
Prostate						1		Lymphoid (first CSO), prostate (second)
Total	4	5	4	5	5	1	4	

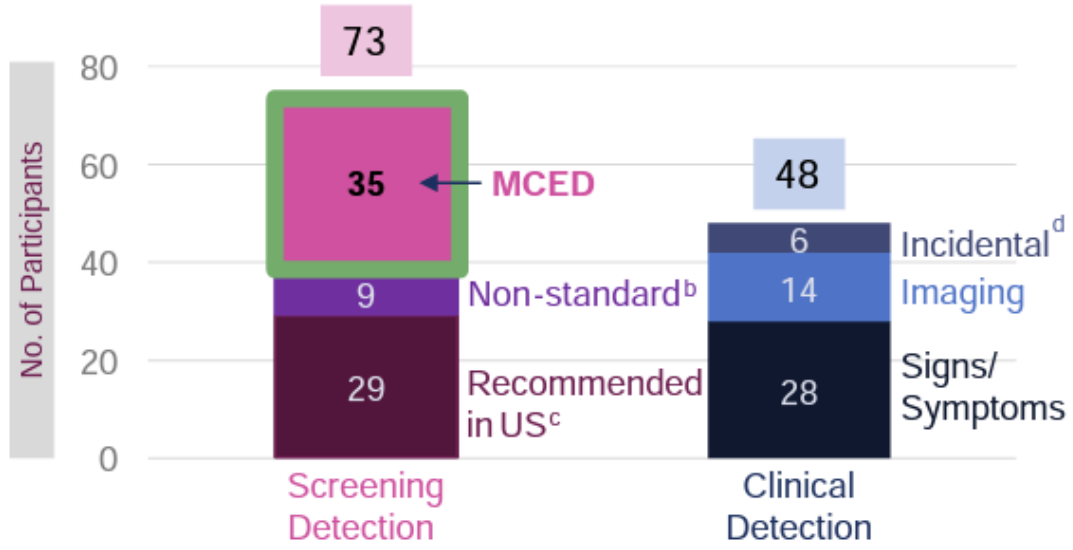
Abbreviations: AJCC, the American Joint Committee on Cancer; CSO, cancer signal origin; GI, gastrointestinal.
[a] Information not available for cancer type/stage/recurrence for one true positive participant at time of analysis.
[b] AJCC version 8.
[c] Unknown stage at time of analysis.
[d] No AJCC stage expected.

Sensitivity under prospective screening
Schrag ESMO 2022; Grail test

Cancers Identified Within One Year of MCED Testing

Participants with Cancers Detected by Either Screening or Clinical Findings

121 participants had a cancer diagnosis within 1 year

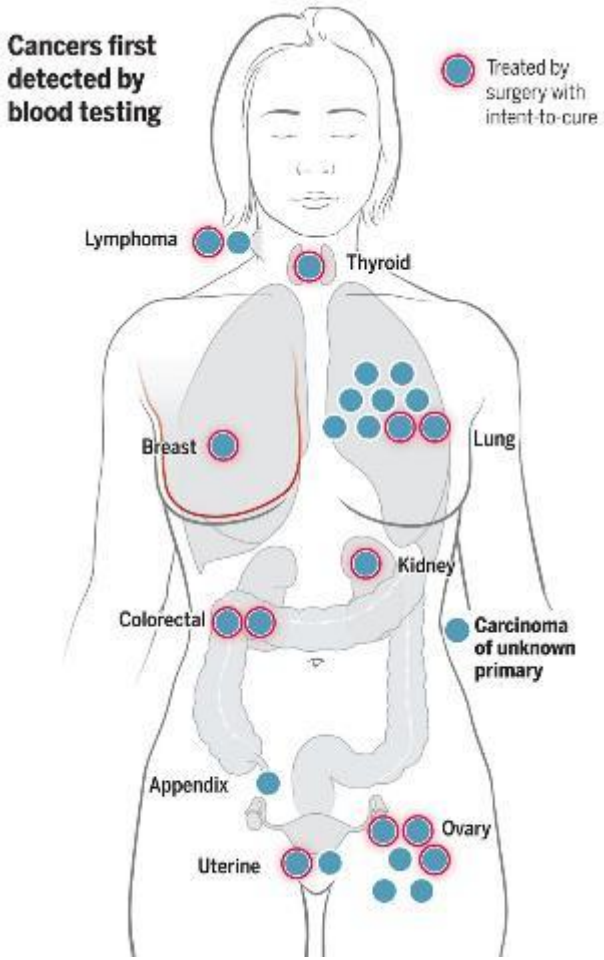


- 35/121 (29%) had cancer diagnosed and positive MCED
  - 2/35 had cancer detected by the MCED test but work-up began before results were disclosed

MCED, multi-cancer early detection.
[a]Based on participants with cancer status assessment at the end of the study.
[b]3 thyroid and 6 melanoma.
[c]Breast, cervical, colorectal, lung, and prostate cancer.
[d]1 incidental radiology finding, 1 incidental finding on routine physical exam, 2 changed lab values, 1 surveillance of prior cancer, 1 follow-up after MGUS diagnosis.

Sensitivity under prospective screening
Lennon et al 2020; EXACT test

- 26/96 (27%) total cancers first identified by MCED



NOTE: Both these sensitivities are proxies for true sensitivity given by

$$\frac{\#\ \textit{screen detected}}{\#\ \textit{screen detected} + \#\ \textit{interval detected}}$$

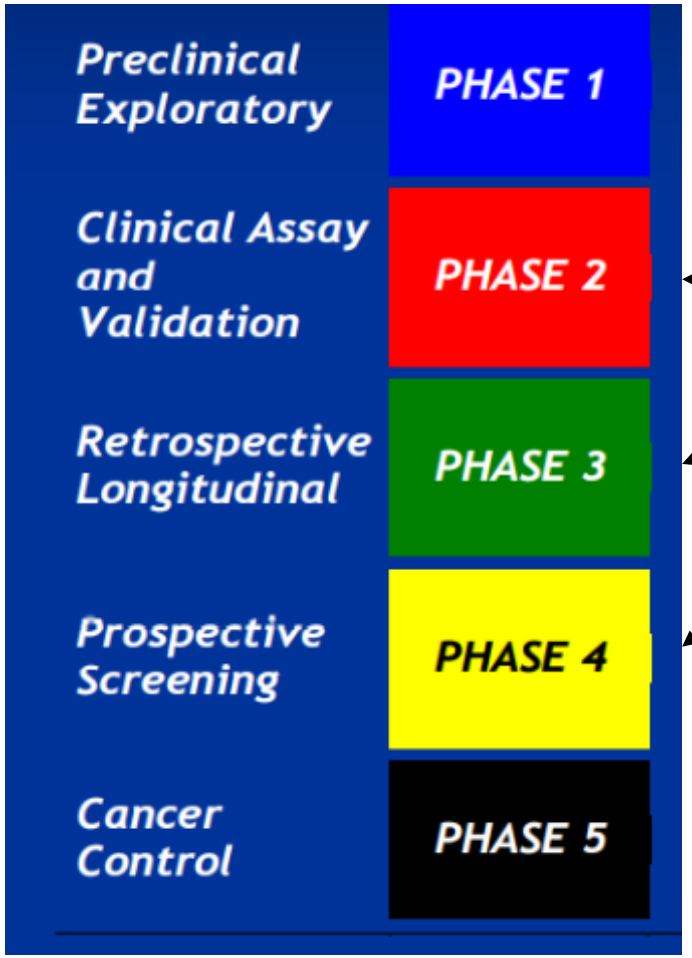
We will refer to this as “empirical sensitivity”

Lennon et al Science 2020

The existing evidence generation pipeline

From diagnostic performance to mortality reduction

NIH NATIONAL CANCER INSTITUTE
Early Detection Research Network



Published studies

National Institutes of Health (.gov)
https://grants.nih.gov › grants › guide › rfa-files › RF...

RFA-CA-23-022: NCI Cancer Screening Research Network

Nov 17, 2022 — A research **network specifically designed to evaluate emerging technologies for cancer screening through randomized controlled trials**, and to ...

Early detection in an evidence crisis

Tests offer the opportunity to potentially detect many more cancers including some that tend to be diagnosed at late stage with high fatality rates

- *Is this likely to provide the promised benefit?*
- *Can we properly evaluate the tests in a timely fashion?*
- *If not, what is the alternative?*
- *Can we leverage understanding of how screening works to develop realistic expectations and generate evidence?*

With screening tests	**Without screening tests**
Breast	Ovary
Prostate	Pancreas
Colorectal	Liver
Cervical	Bladder
Lung	Stomach

The need for evidence about outcomes

An ovarian cancer case study

- Novel blood-based biomarker CA125 – early 1980s
- Algorithmic test (ROCA) thresholded to produce high specificity
  - ROCA uses longitudinal behavior of CA125 to predict presence of cancer
- Empirical sensitivity in UKCTOCS screening trial: 85%
  - For MMS screening; ROCA with triage to ultrasound
- Trial outcomes
  - Non-significant 15% mortality reduction in first report
  - No reduction in mortality at second report (after stop screen interval)
  - Despite 24% drop in incidence of stage IV disease
- Are these findings due to the trial, the test, or the cancer?

Jacobs et al Lancet 2016; Menon et al Lancet 2021

The ingredients of screening benefit

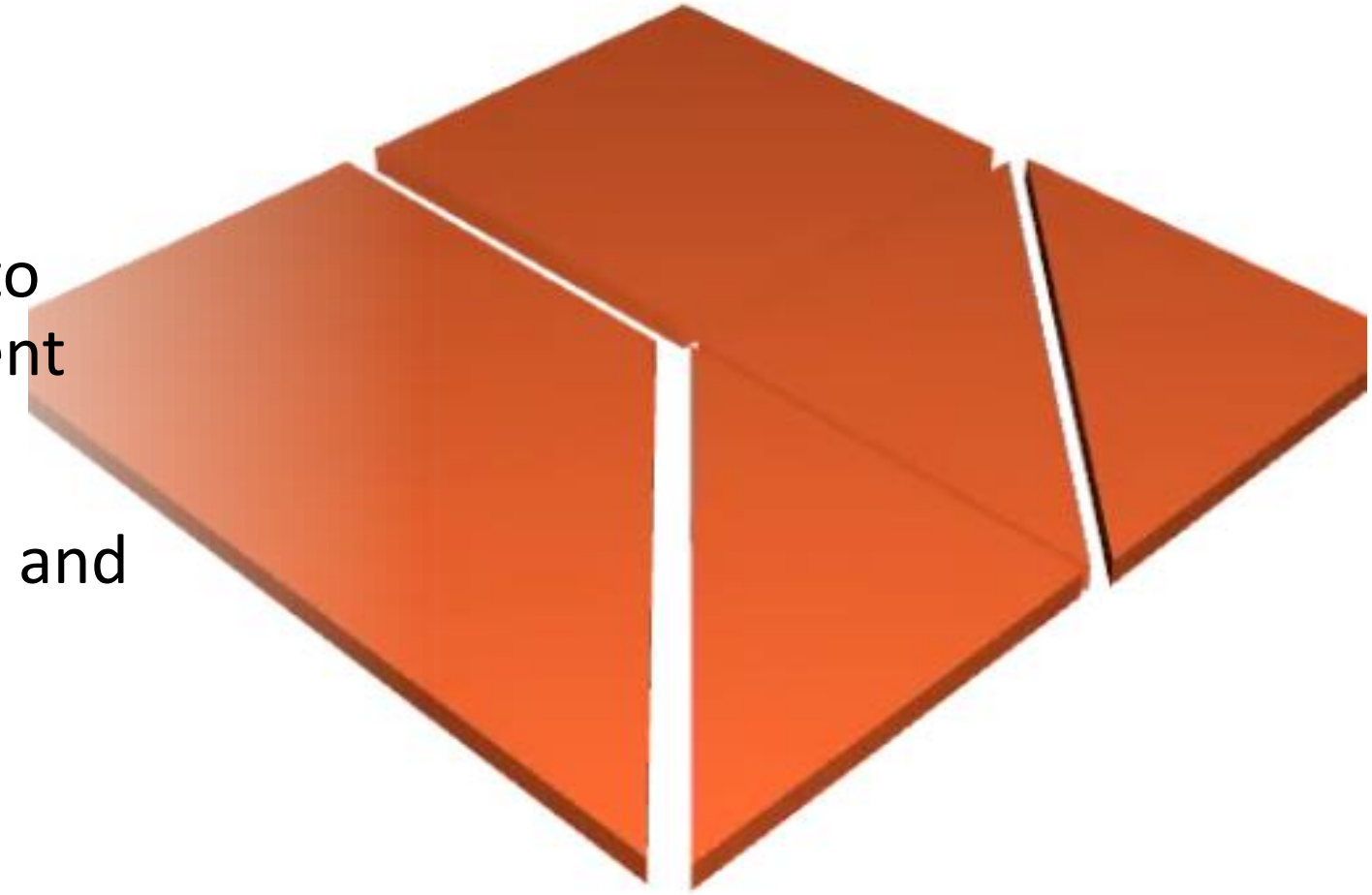
1. **Sensitivity to detect disease early**
2. **Opportunity to detect and intercept disease early**
3. **Translation from early* detection to mortality reduction**

Early*: disease shifted earlier by screening

An evidence generation proposal

1. Develop studies/methods to inform about each ingredient

2. Combine them via rigorous and transparent models



Sensitivity of a test

Likelihood a test will be positive if the cancer is there

Different versions of sensitivity

A. Sensitivity to detect cancer in known cases

Established first and common in early studies of test performance

B. Sensitivity to detect cancer before clinical diagnosis

Much harder to assess but key driver of benefit

Sensitivity A versus Sensitivity B

Sensitivity A: sensitivity among known cases

- Known cases already diagnosed by existing means
- Stage mix by design or based on sample availability

Sensitivity B: sensitivity in intended-use population

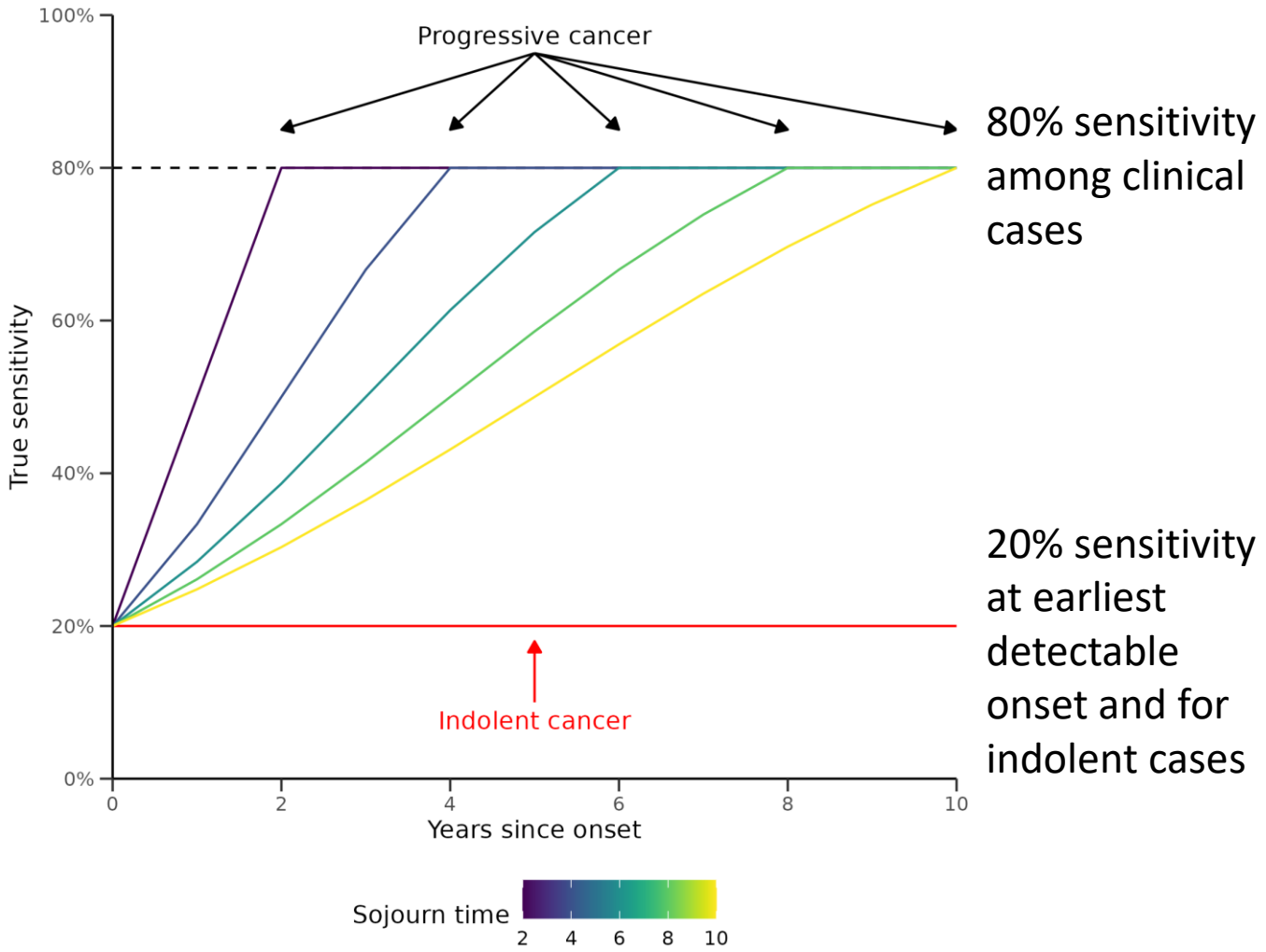
- Timing in cases is earlier than when they would have been diagnosed
- Stage distribution likely skewed towards early stages
- Actual driver of benefit is **screening episode sensitivity**
- A common proxy

Empirical sensitivity = $\dfrac{\#\ \textit{screen detected}}{\#\ \textit{screen detected} + \#\ \textit{interval detected}}$

From Sensitivity A to Sensitivity B

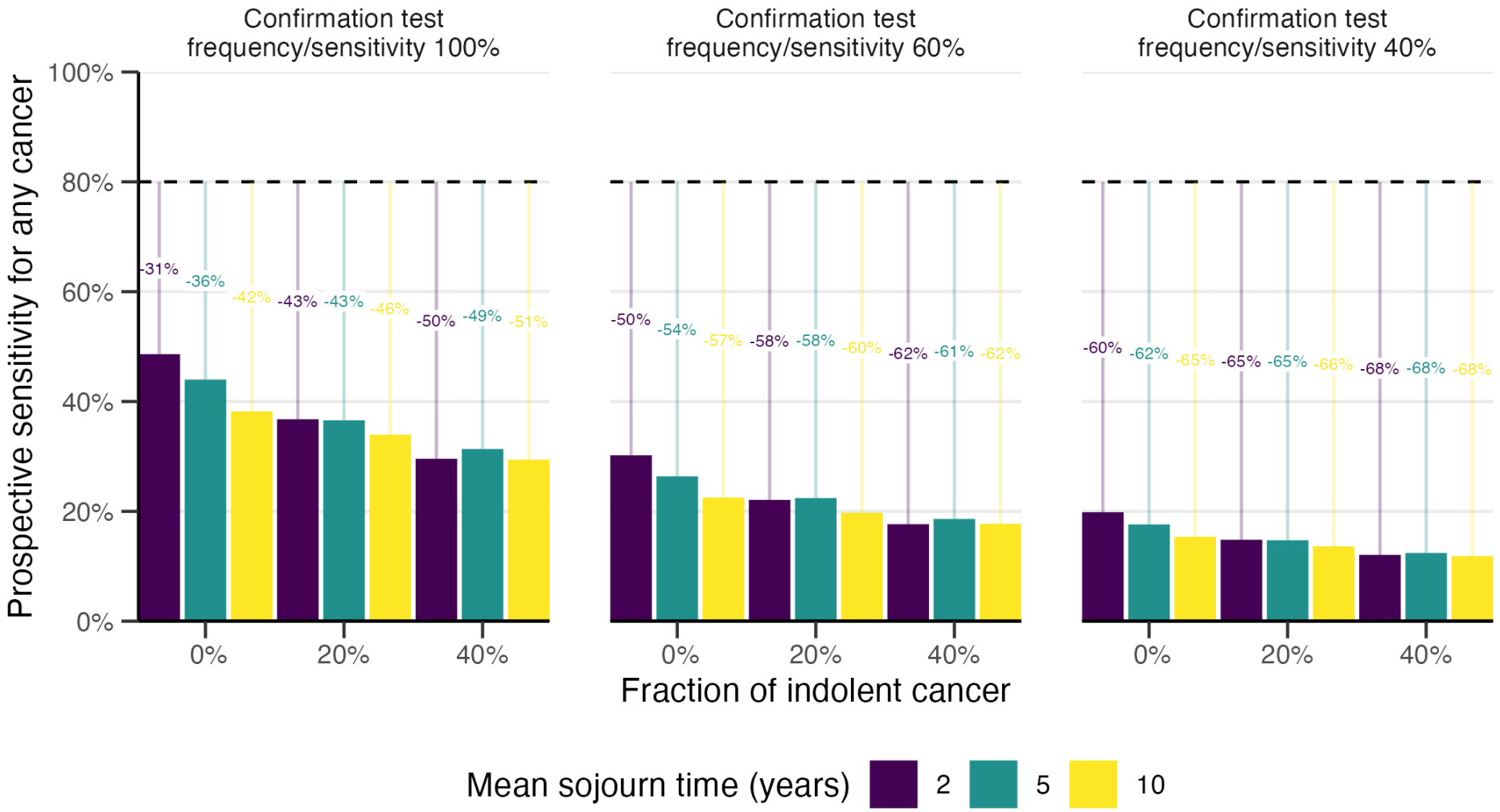
Realistic degradation of performance

1. Specify a sensitivity curve
2. Specify mean sojourn time
3. Specify fraction non-progressive w/ low sensitivity
4. Specify access to and accuracy of confirmation testing



Zhao et al unpublished

From Sensitivity A to Episode Sensitivity



Estimating Cancer Screening Sensitivity and Specificity Using Healthcare Utilization Data: Defining the Accuracy Assessment Interval

CEBP August 2022

Jessica Chubak[1,2], Andrea N. Burnett-Hartman[3,4], William E. Barlow[5], Douglas A. Corley[6], Jennifer M. Croswell[7], Christine Neslund-Dudas[8], Anil Vachani[9], Michelle I. Silver[10], Jasmin A. Tiro[11,12], and Aruna Kamineni[1]

Recent papers addressing bias of empirical sensitivity

Test sensitivity in a prospective cancer screening program: A critique of a common proxy measure

SMMR 2023

Jane Lange[1], Yibai Zhao[2], Kemal Caglar Gogebakan[1], Antonio Olivas-Martinez[3], Marc D. Ryser[4], Charlotte C. Gard[5], and Ruth Etzioni[1,3,6]

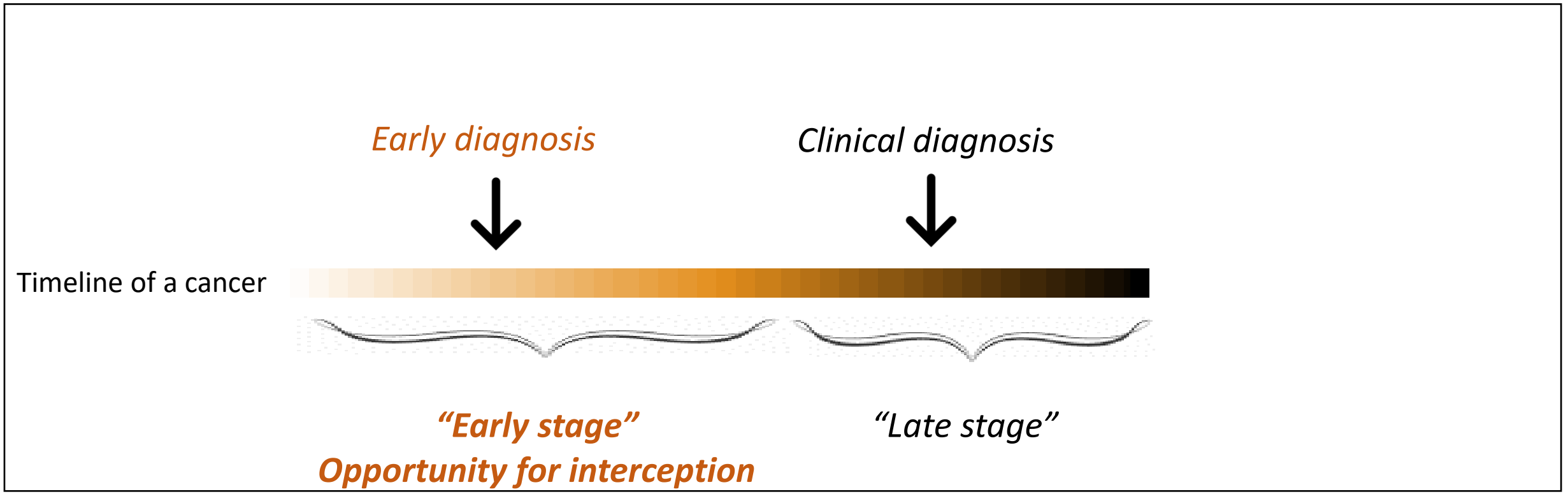
The ingredients of screening benefit

1. Sensitivity to detect disease early
2. **Opportunity to detect and intercept disease early**
3. **Translation from early* detection to mortality reduction**

Early*: disease shifted earlier by screening

Opportunity for interception

Among cases that would otherwise progress to late stage

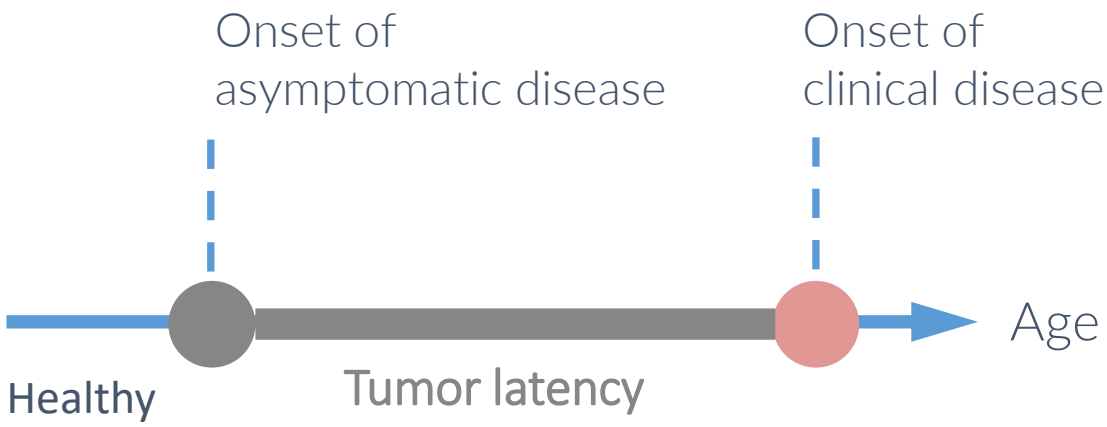


Interval during which cancer detectable by existing diagnostic means and its fate potentially changeable by existing treatments

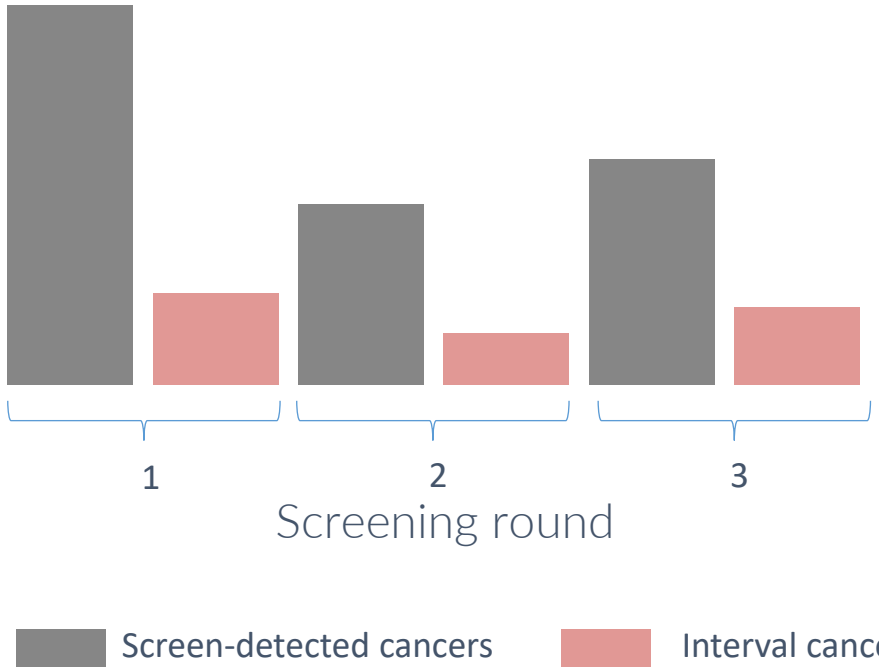
How do we learn about opportunity?

From studying changes in disease incidence under screening

Disease natural history

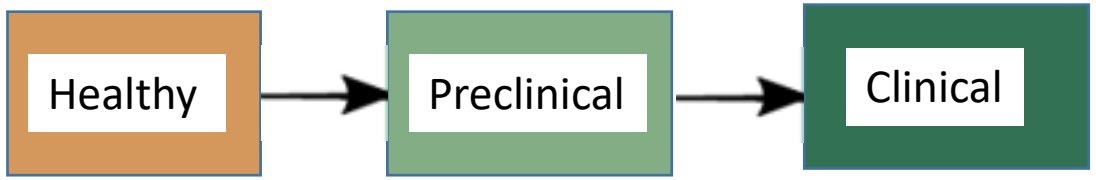


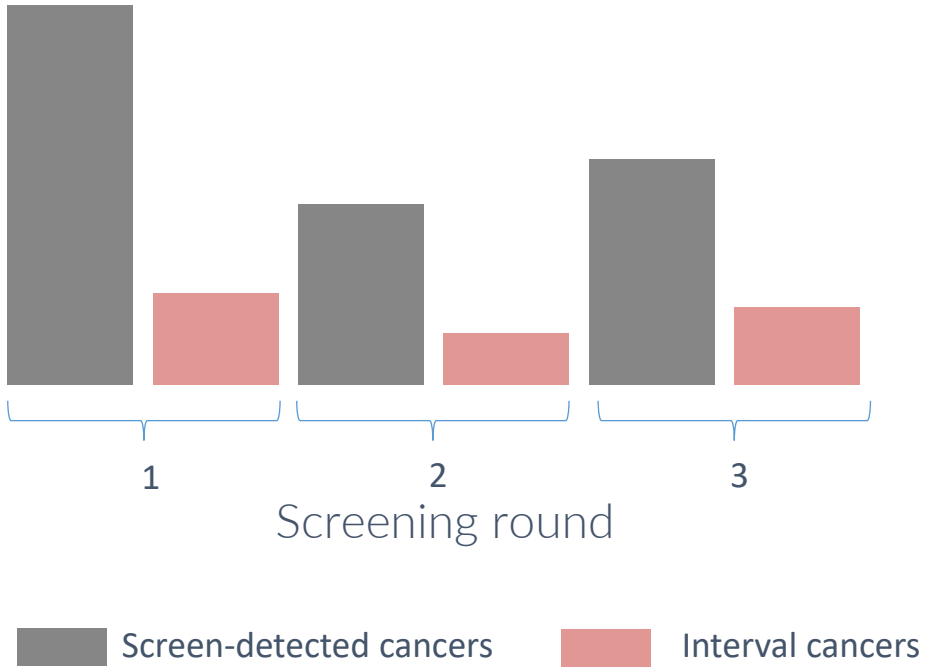
Patterns of cancer detection



How do we learn about opportunity?
From studying changes in disease incidence under screening

Model of disease progression ← Patterns of cancer detection

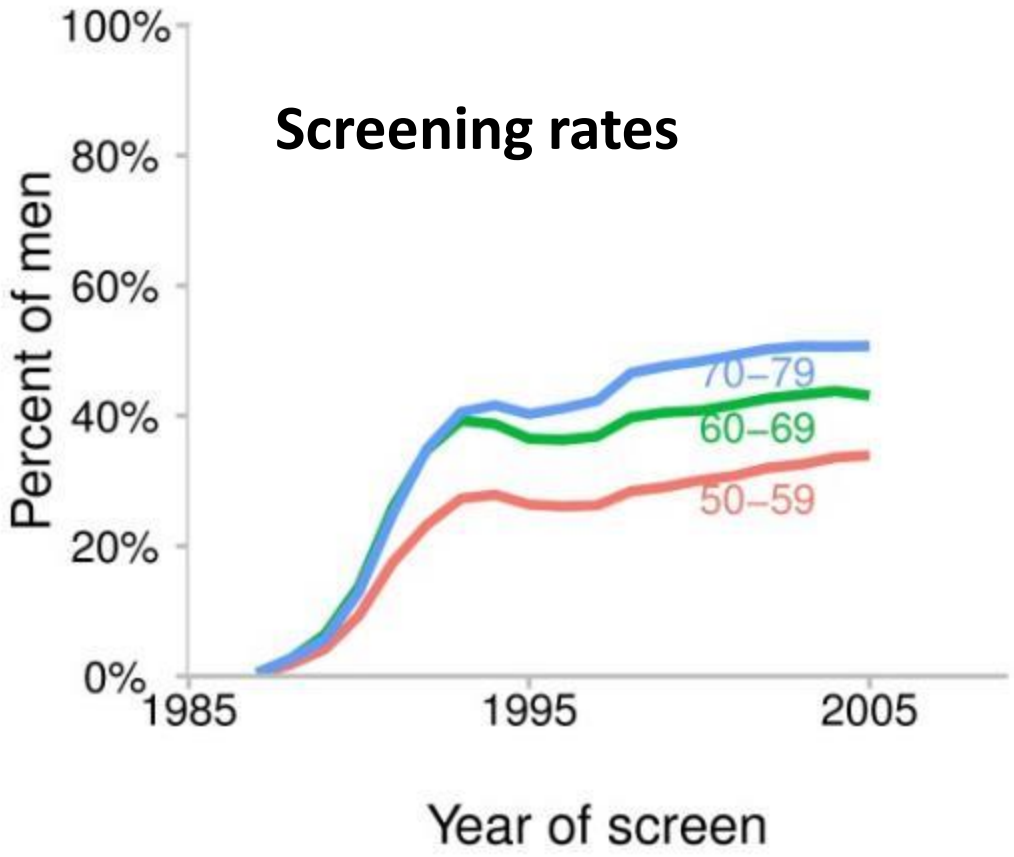


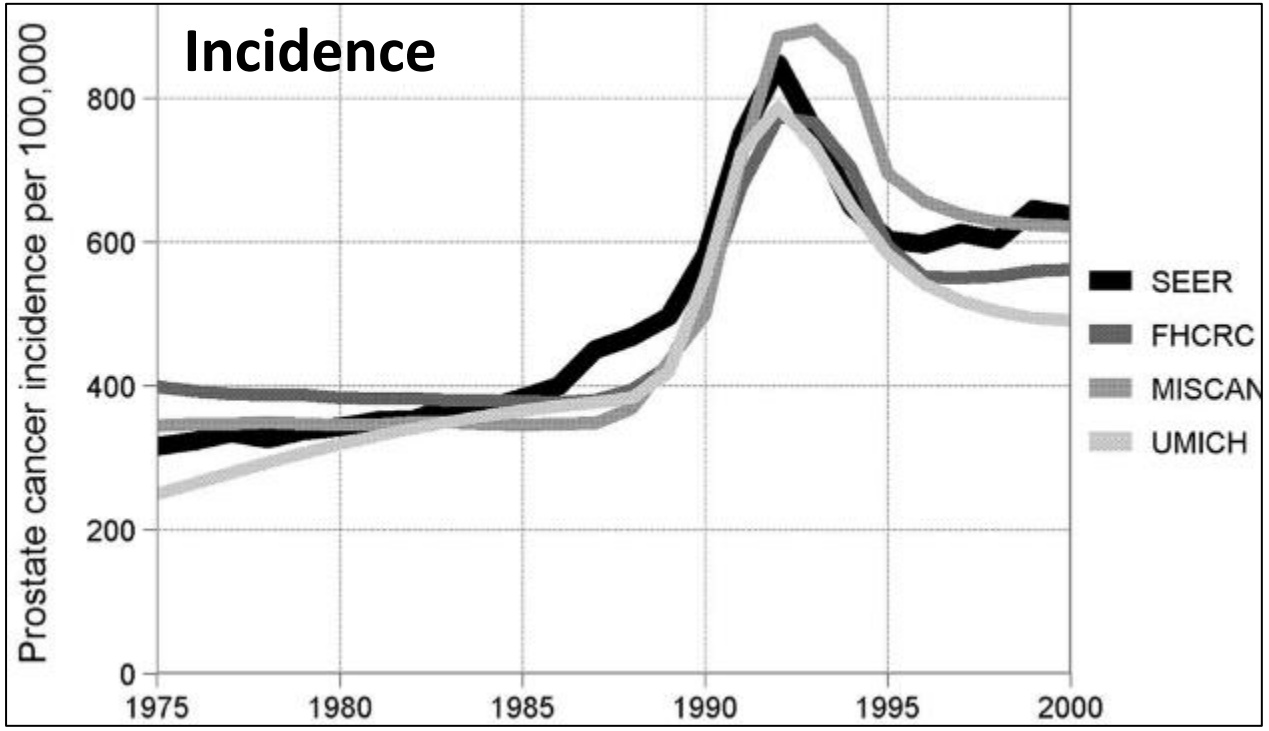


- Learn transition rates between disease states
- Results will vary depending on test sensitivity

Shen and Zelen JCO 2001; Ryser et al Ann Int Med 2022

Preclinical duration in prostate cancer
Learned from US incidence trends (SEER)





Mean latency
7-14 years

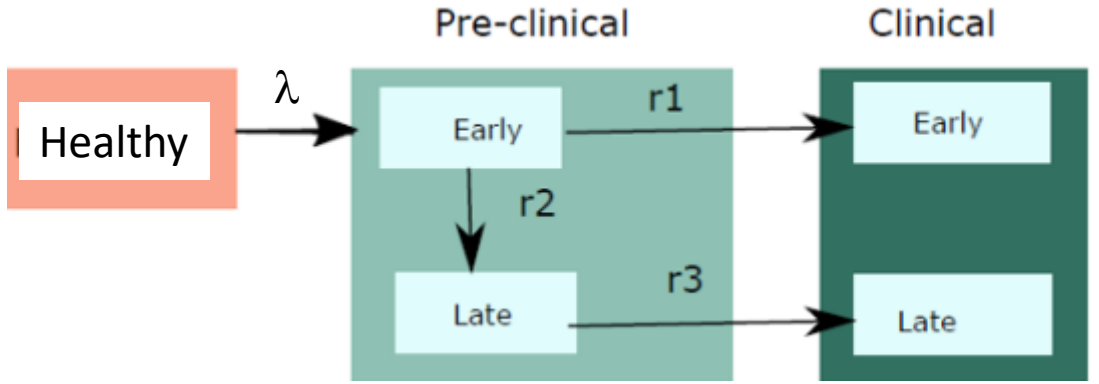
Gulati et al CEBP 2011

Latencies learned for different cancers

Prostate *Gulati et al CEBP 2011*	7-14 years	Different estimation methods US population data
Colorectal *Rutter et al Med Dec Making 2016*	3.5-5 years	Different estimation models Combination of data sources
Lung *Ten Haaf et al CEBP 2015*	4 years	One model/estimation method Data from PLCO/NLST
Breast *Ryser et al Annals Int Med 2022*	3.5-6.5 years	Different methods, calendar periods Screening trials and BCSC data

Learning the early-stage latency

More challenging than learning overall latency



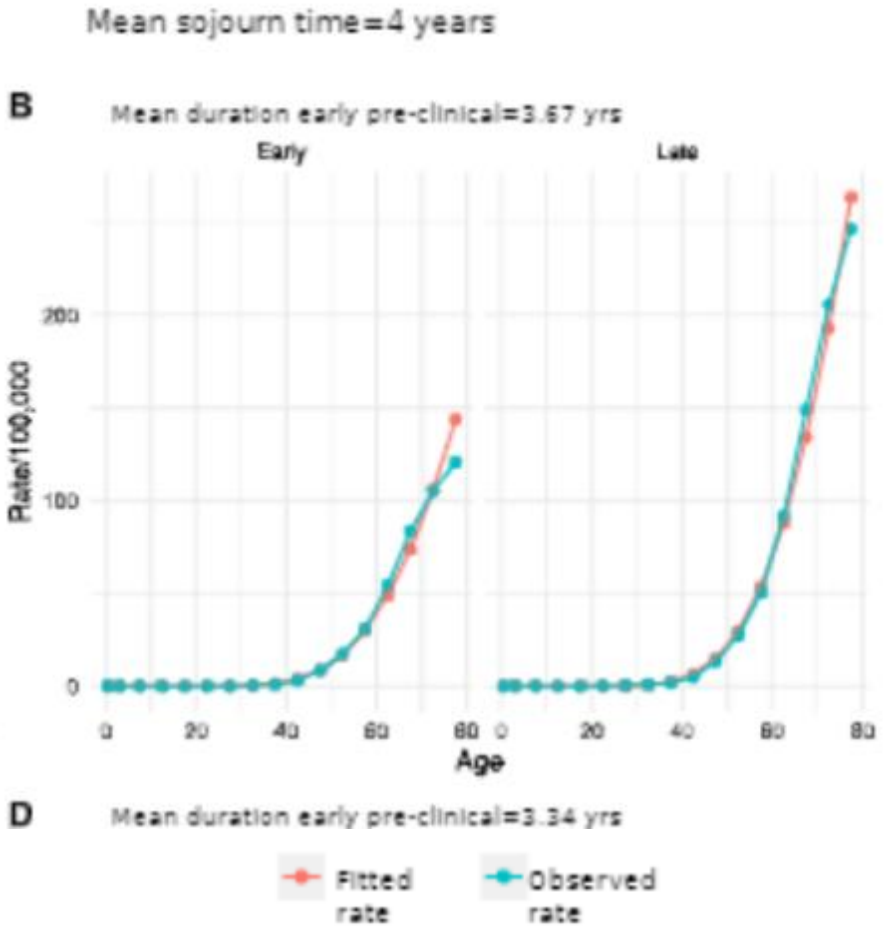
- Requires high-quality data on
  - Screen- and interval cancer incidence
  - Disease stage at diagnosis
  - Screening and biopsy frequencies
  - An estimate of screening test sensitivity B ideally by stage
- In cancers with screening programs we may have this information but learning early-stage latency is still hard
- In cancers without screening trials or programs we lack the needed data and must add information to fill in the gaps

Pinsky Biometrics 2004

Adding information
One approach - lung cancer

- Add information on:
  - **Mean overall preclinical latency (OMST)**
  - **Preclinical late-stage latency (LMST)**
- Permits estimating early-stage duration from SEER age- and stage-specific incidence
- Superimpose any screening protocol with specified sensitivity to project impact on late-stage incidence

Mean overall latency 4y
Mean preclinical late stage 0.5 y
Mean early-stage latency 3.67 y



SEER age-specific incidence for early and late stage

Lange et al under review

Results for three cancers

Annual screening for 3 years with 30% early-stage sensitivity

Early Stage Mean Sojourn Time (EMST) by Cancer Site, OMST and LMST

OMST: Overall mean sojourn time LMST: Time in late stage before clinical detection

	OMST=3.5 LMST=0.5	OMST=3.5 LMST=1	OMST=2 LMST=0.5	OMST=2 LMST=1
Liver	3.29	3.07	1.79	1.58
Pancreas	3.04	2.59	1.54	1.09
Bladder	3.43	3.37	1.93	1.87

Late-stage incidence reduction
Assumes full access to accurate diagnostic confirmation

29-30% 23-25% 22-25% 14-20%

Lange et al 2023 under review

Pattern of late-stage incidence reduction

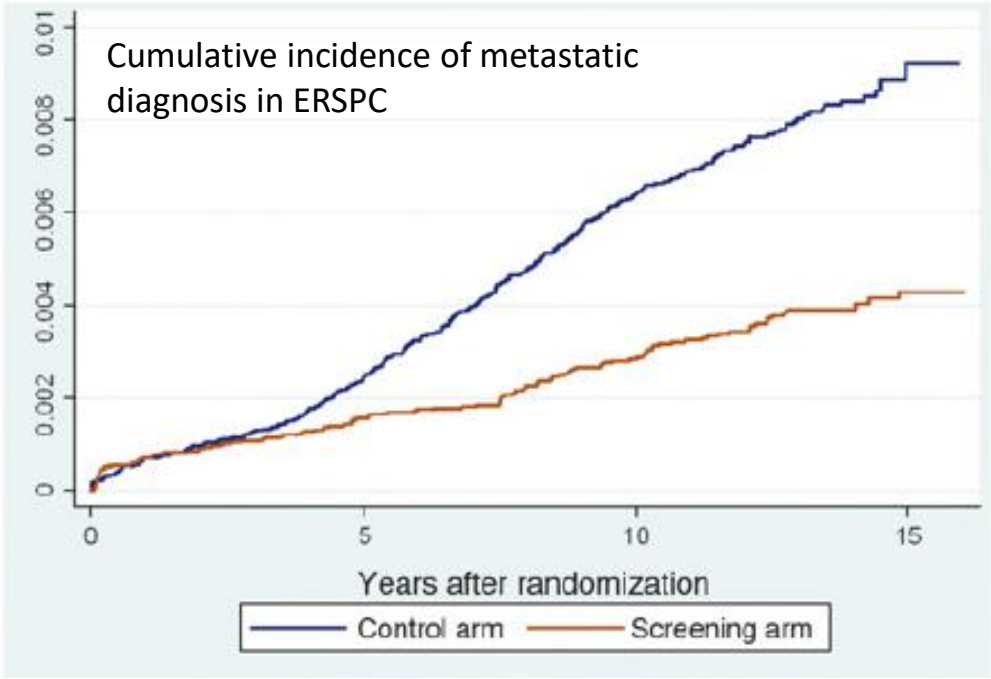
- Longer duration of preclinical late stage
  - Takes longer to see drop in late-stage incidence
  - May lead to initial excess of late-stage cases in screen arm
- Longer duration of preclinical early stage
  - More pronounced late-stage drop eventually
  - May take longer to see any drop
- Pattern of late-stage incidence in a trial may be informative about stage durations

Platinum Priority – Prostate Cancer

Editorial by Allison S. Glass, Matthew R. Cooperberg and Peter R. Carroll on pp. 753–755 of this issue

Screening for Prostate Cancer Decreases the Risk of Developing Metastatic Disease: Findings from the European Randomized Study of Screening for Prostate Cancer (ERSPC)

(b) Risk ratio 0.503 (0.406–0.622)



Schroder et al Eur Urol 2017

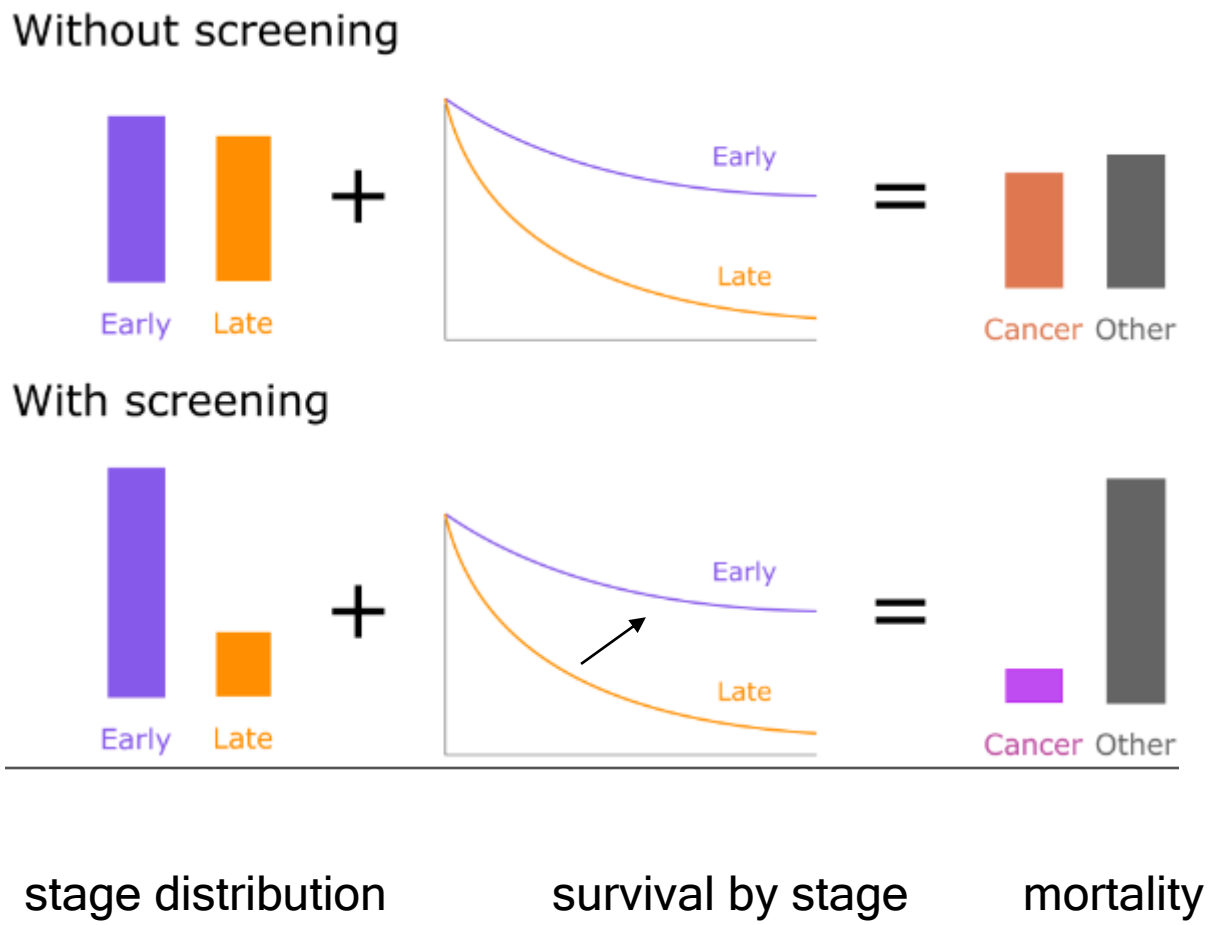
The ingredients of screening benefit

1. Sensitivity to detect disease early
2. Opportunity to detect and intercept disease early
3. **Translation from early* detection to mortality reduction**

Early*: disease shifted earlier by screening

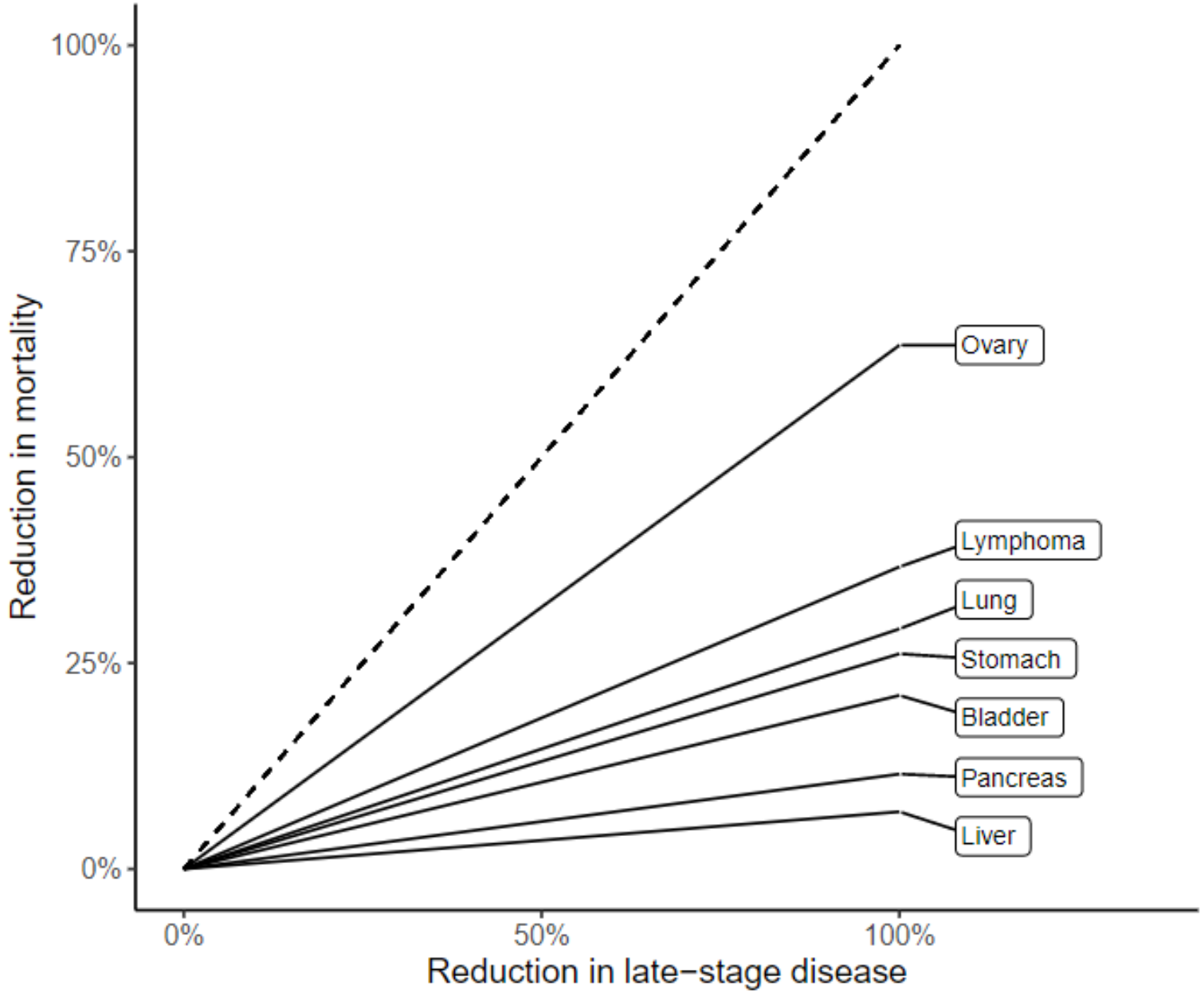
From reduction in late stage to reduction in mortality

- If late-stage incidence is reduced by $\propto$ % what should we expect regarding mortality reduction?
- The most common approach is to assume that the curability of a case shifted earlier by screening is driven by their new stage
- Predict reduction in mortality by replacing disease survival for cases shifted out of late stage by the survival of early- stage cases
- Under some fairly basic assumptions the predicted reduction in mortality is proportional to but less than $\propto$ %



Relative mortality reduction as a function of ∝ for different cancers in SEER

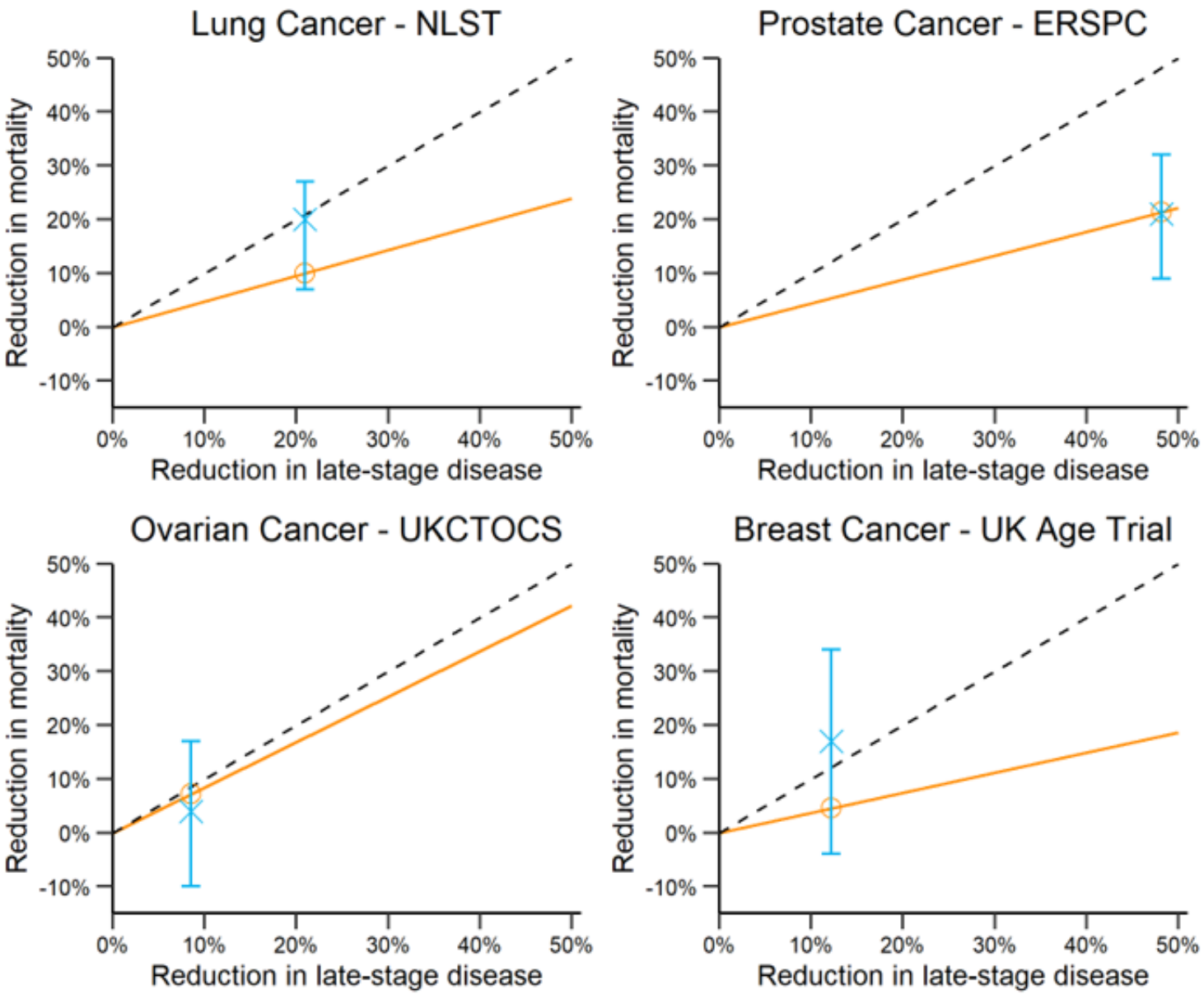
For a given reduction in late stage ∝ , variable predicted reduction in mortality across cancers



Owens L et al CEBP 2022

Modeled and observed mortality reduction in four published trials given observed ∝

x observed
o predicted



Owens et al CEBP 2022

Towards a better prediction of the effect of screening on mortality – incorporating subtypes

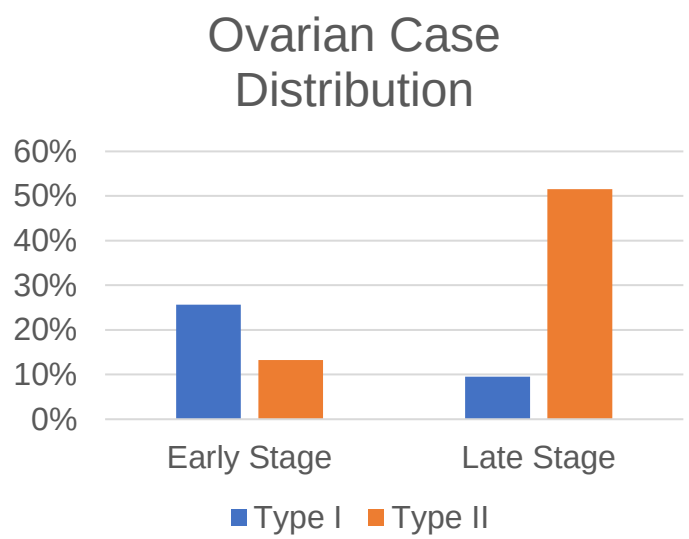
- Any cancer consists of a mix of prognostic subtypes

Ovarian cancer	Type I and Type II
Breast cancer	ER+ and ER–
Prostate cancer	Gleason Score 6 or 7+

- Distributions of subtypes typically differ for early versus late stage
- Simple predictions of mortality based on substituting early- for late-stage survival may be unwittingly altering the subtype distribution

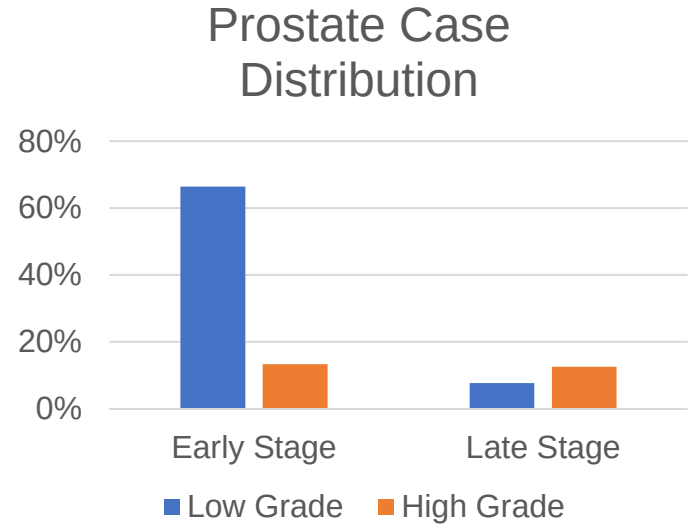
Owens et al CEBP 2022, Owens et al, CEBP 2023

Ovarian and prostate cancer by subtype and stage



Ovarian cancer consists of two histological subtypes

- Type I – *less aggressive, rarely diagnosed in late stage*
- Type II – *more aggressive, more common in late stage*



Prostate cancer consists of low- and high-grade cases

- Low grade – *less aggressive, rarely diagnosed in late stage*
- High grade – *more aggressive, more common in late stage*

Ovarian Cancer Mortality Reduction With and Without Subtype: 30% late-stage reduction proportionate to type in late stage			
	Type I	Type II	Total
Stage Shift Multiplier	0.61	0.60	0.69
Projected Mortality Reduction			
Ignoring Type			**21%**
Preserving Type			**18%**

Prostate Cancer Mortality Reduction With and Without Subtype: 30% late-stage reduction proportionate to type in late stage			
	Gleason 6-	Gleason 7+	Total
Stage Shift Multiplier	0.40	0.29	0.44
Projected Mortality Reduction			
Ignoring Type			**13%**
Preserving Type			**10%**

Owens et al CEBP 2023

Results hold implications for late-stage incidence reduction as a "surrogate" endpoint

Multi-cancer early detection technologies: a review informed by past cancer screening studies

The future of cancer screening and lessons from the past

Raoof et al CEBP 2021

"Randomized control trials (RCTs) to show mortality reduction have required millions of screening-years, two-decade durations, and been susceptible to external confounding. Future RCTs with ***late-stage incidence as a surrogate endpoint*** *could substantially reduce these challenges"*

Predicted mortality reduction vs late-stage incidence reduction as endpoint

J. R. Statist. Soc. A (1996)
159, *Part* 1, *pp.* 49–60

Trial Design Based on Surrogate End Points — Application to Comparison of Different Breast Screening Frequencies

By N. E. DAY and S. W. DUFFY†

Surrogate endpoints for cancer screening trials: general principles and an illustration using the UK Flexible Sigmoidoscopy Screening Trial

Jack Cuzick, Fay H Cafferty, Robert Edwards, Henrik Møller and Stephen W Duffy

J Med Screen 2007;**14**:178–185

Review and Summary

The field of cancer early detection is heading for an evidence crisis

- Screening trials with mortality endpoint ideal but may not be timely enough
- Understanding drivers of benefit will be key in evidence generation
  - We expect prospective diagnostic performance to be considerably degraded
  - We may still see a stage shift in short-term trials depending on setting and cancers
  - In absence of within-stage shifts or improvements in diagnostic imaging and early-stage treatments expect a modest impact on mortality
- Integrate models with observational and judiciously conducted clinical studies
- Need a parallel effort to quantify harms and system impacts

Forthcoming commentary (JNCI); Revisiting the Standard Blueprint for Biomarker Development to Address Emerging Cancer Early Detection Technologies. Etzioni et al 2023

Team and support

- Lukas Owens
- Jane Lange
- Roman Gulati
- Yibai Zhao
- Noel Weiss
- Boshen Jiao
- Kemal Gogebakan

Rosalie and Harold Rea Brown chair at Fred Hutch

CEDAR at the Knight Cancer Institute

NCI's Cancer Intervention and Surveillance Modeling Network

NCI R35 Modeling and Analytics for Novell Cancer Diagnostics

Email: retzioni@fredhutch.org

http://mced-calculator.fredhutch.org *for our multi-cancer test calculator that permits configuration of a multi-cancer test and projection of selected outcomes*

https://lukasowens.shinyapps.io/stage-shift/ *for stage shift model predictions*

https://cedarmodelingframework.shinyapps.io/mced_app/ *for stage projections*