UK-China Health and Economy Partnership
中英健康医疗与经济合作项目

BRITISH COUNCIL

gsk

Bournemouth University

FUDAN UNIVERSITY 1905

UNIVERSITY of York

SHANDONG UNIVERSITY 1901

The University Of Sheffield.

ZHEJIANG UNIVERSITY 1897

UNIVERSITY OF LEEDS

Thinking Beyond Accuracy

Dr Bethany Shinkins

Associate Professor, University of Leeds, UK.

b.shinkins@leeds.ac.uk

Learning Objectives

Accuracy is a key component of test evaluation, but we are mainly interested in the impact that tests have on patient outcomes (clinical utility)

Modelling the downstream consequences of tests is essential to capture clinical utility

Linking decision trees to Markov models is a common technique used to link test results to their impact on long-term costs and outcomes

Over to you

Let's say the diagnostic accuracy of a new test has been evaluated and demonstrated to be highly accurate

How may this test still fail to improve patient outcomes?

An accurate test may not improve outcomes

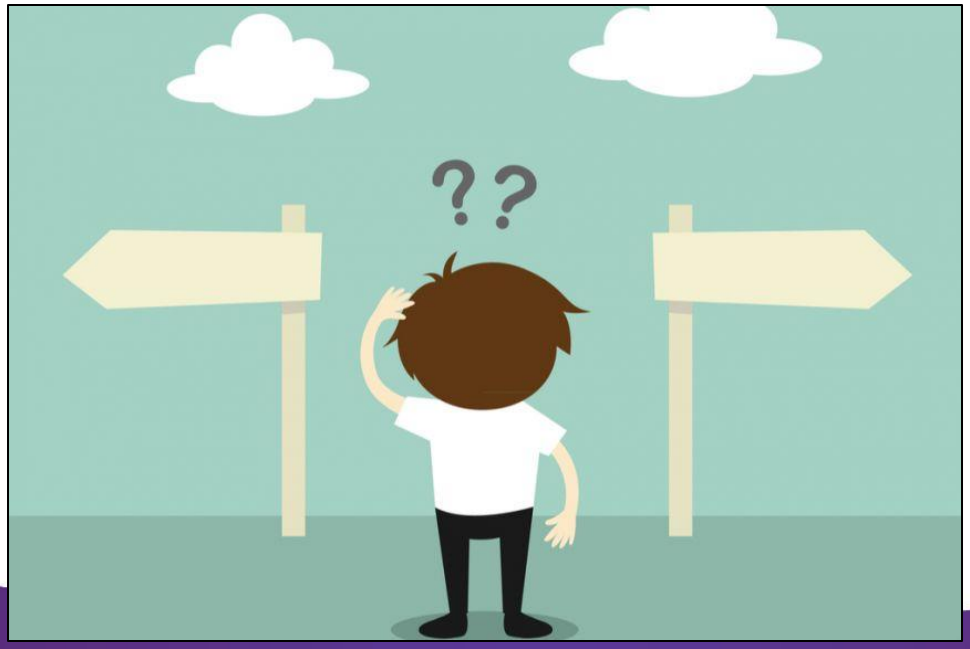
- Test result is too slow to change management
- Test result doesn't make it back to the ordering physician
- Patient is already too sick/too healthy for the test result to matter
- Test is performed inappropriately
- Result of test is acted upon inappropriately
- The test in question is only one of many tests ordered
- Treatment is not available (too expensive, out of stock, etc.)
- Treatment is not delivered
- Patient declines the treatment

….And the list goes on!

Accuracy just one part of the picture

Clinical utility hinges on whether patients and clinicians use these interventions and how they respond to test results (e.g. whether patients undergo appropriate treatment)

Buchanan et al. 2013

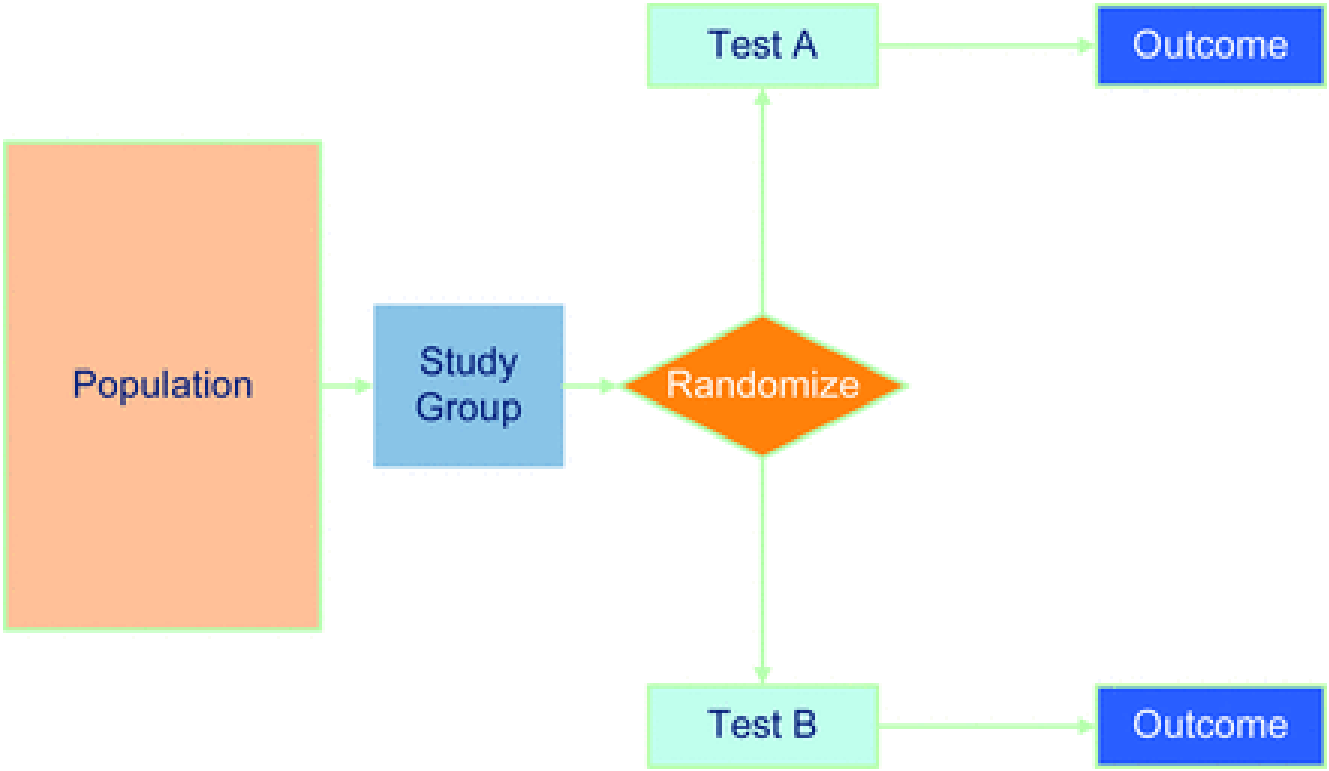


Generating evidence on the downstream consequences of testing on patient outcomes and costs is key to developing a convincing cost-effectiveness analysis

Two main methods:

1) Direct comparison within a randomised controlled trial

2) Decision analytic models that link together multiple sources of evidence



Clin Chem, Volume 58, Issue 12, 1 December 2012, Pages 1636–1643, https://doi.org/10.1373/clinchem.2012.182576

- Test-treatment RCTs are rare
  - Average of 37 published per year (Ferrante di Ruffano, 2012)

- Long follow-up often required
  - Beyond the length of a feasible study
  - Technology obsolete by the time evaluation complete?

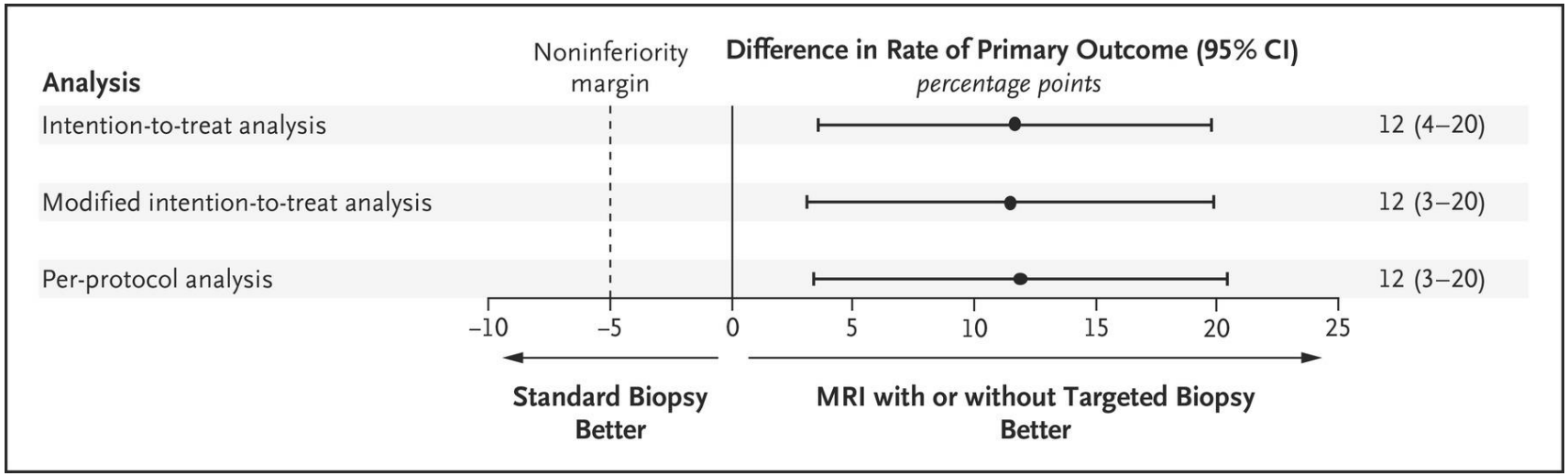
- RCTs evaluate consequences of test but also full patient management strategy (Bossuyt, 2012)
  - Average test could improve outcomes and cost-effectiveness if management effective
  - Good test could fail to improve outcomes and be cost-effective if management ineffective

Example: the PRECISION study

- Non-inferiority RCT
- MpMRI with targeted biopsy vs. standard transrectal US–guided biopsy for detection of clinically significant prostate cancer
- Proportion of men who received a diagnosis of clinically significant cancer
- Patients have agreed for their longer-term (e.g. mortality) to be followed up (ongoing)

Kasivisvanathan et al., NEJM, 2018

Example: the PRECISION study



Kasivisvanathan et al., NEJM, 2018

THE LANCET

ARTICLES | VOLUME 389, ISSUE 10071, P815-822, FEBRUARY 25, 2017

Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study

Hashim U Ahmed, FRCS * • Ahmed El-Shater Bosaily, MBBCh * • Louise C Brown, PhD * • Rhian Gabe, PhD • Prof Richard Kaplan, FRCP • Prof Mahesh K Parmar, DPhil • et al. Show all authors • Show footnotes

Open Access • Published: January 19, 2017 • DOI: https://doi.org/10.1016/S0140-6736(16)32401-1 •

The PROMIS study:

- Compared the diagnostic accuracy of MP-MRI and TRUS-biopsy against a reference test (TPM-biopsy)
- Men with no previous biopsy (n=740), underwent MP-MRI followed by both TRUS-biopsy and TPM-biopsy. The conduct and reporting of each test was done blind to other test results.
- MP-MRI was more sensitive than TRUS-biopsy (93% vs 48%) and less specific (41% *vs* 96% TRUS-biopsy).
- Collected data on procedure-related adverse events: 44 (5·9%) of 740 patients reported serious adverse events, including 8 cases of sepsis.

Ahmed et al. 2017. The Lancet

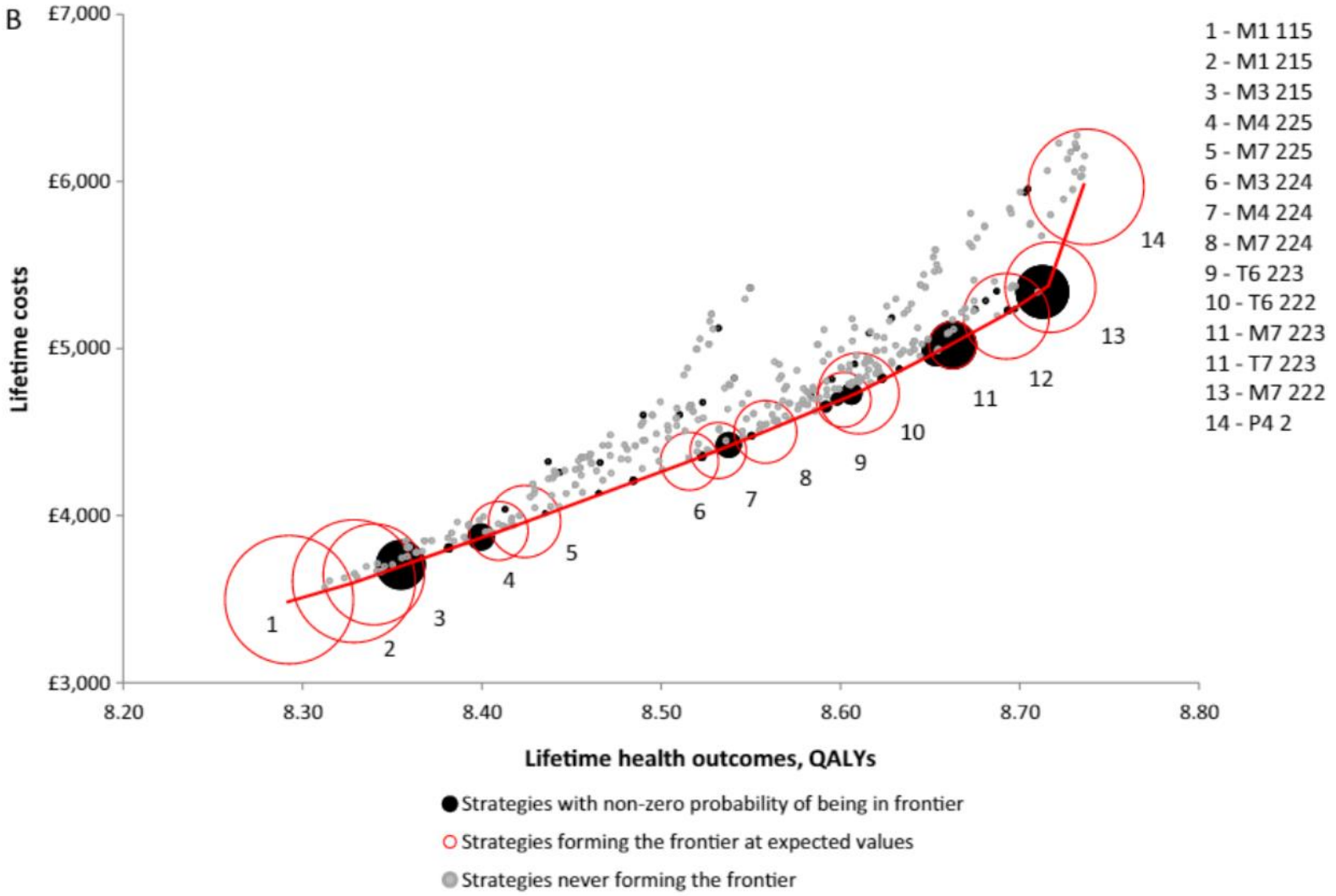
- Decision models link evidence from different types of studies together

- Data on intermediate outcomes (such as accuracy) are linked (by making certain assumptions) to data on long-term outcomes

Faria et al. (2018) used a linked evidence approach to build a decision analytic model to evaluate the cost-effectiveness of different diagnostic strategies for prostate cancer, using the data from the PROMIS study

Decision tree for immediate diagnostic pathway, Markov cohort model for longer term costs and outcomes

There were 32 different test combinations, and possibility of using different cut-offs and classifications of disease: resulted in comparison of 383 different diagnostic strategies



Faria et al. (2018) European Urology

NICE guidance updated

Magnetic resonance imaging and biopsy

1.2.1 Do not routinely offer multiparametric MRI to people with prostate cancer who are not going to be able to have radical treatment. **[2019]**

1.2.2 Offer multiparametric MRI as the first-line investigation for people with suspected clinically localised prostate cancer. Report the results using a 5-point Likert scale. **[2019]**

1.2.3 Offer multiparametric MRI-influenced prostate biopsy to people whose Likert score is 3 or more. **[2019]**

1.2.4 Consider omitting a prostate biopsy for people whose multiparametric MRI Likert score is 1 or 2, but only after discussing the risks and benefits with the person and reaching a shared decision (see table 1). If a person opts to have a biopsy, offer systematic prostate biopsy. **[2019]**

https://www.nice.org.uk/guidance/ng131

VALUE IN HEALTH 17 (2014) 22–33

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

The Cost-Effectiveness of a Pharmacogenetic Test: A Trial-Based Evaluation of *TPMT* Genotyping for Azathioprine

Alexander J. Thompson, MSc[1], William G. Newman, FRCP, PhD[2], Rachel A. Elliott, PhD, BPharm, MRPharmS[3], Stephen A. Roberts, PhD, BSc[1], Karen Tricker, PhD, MPM[2], Katherine Payne, PhD, MSc, BPharm, MRPharmS[1,]*

Payne et al. [18] previously concluded that TPMT testing, using either a genotype-based test or a phenotype-based test, was a cost-effective use of resources. This statement, however, was based on a systematic review of published model–based economic evaluations

VALUE IN HEALTH 17 (2014) 22–33

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

The Cost-Effectiveness of a Pharmacogenetic Test: A Trial-Based Evaluation of *TPMT* Genotyping for Azathioprine

Alexander J. Thompson, MSc[1], William G. Newman, FRCP, PhD[2], Rachel A. Elliott, PhD, BPharm, MRPharmS[3], Stephen A. Roberts, PhD, BSc[1], Karen Tricker, PhD, MPM[2], Katherine Payne, PhD, MSc, BPharm, MRPharmS[1,]*

All these previous models assumed that clinicians would completely adhere to the recommended dosing strategies. The present study reaches a more cautious conclusion, because clinicians were not following the test recommendations, which in effect diluted the potential added value of using the TPMT genotyping test.

VOLUME 28 · NUMBER 10 · APRIL 1 2010

JOURNAL OF CLINICAL ONCOLOGY

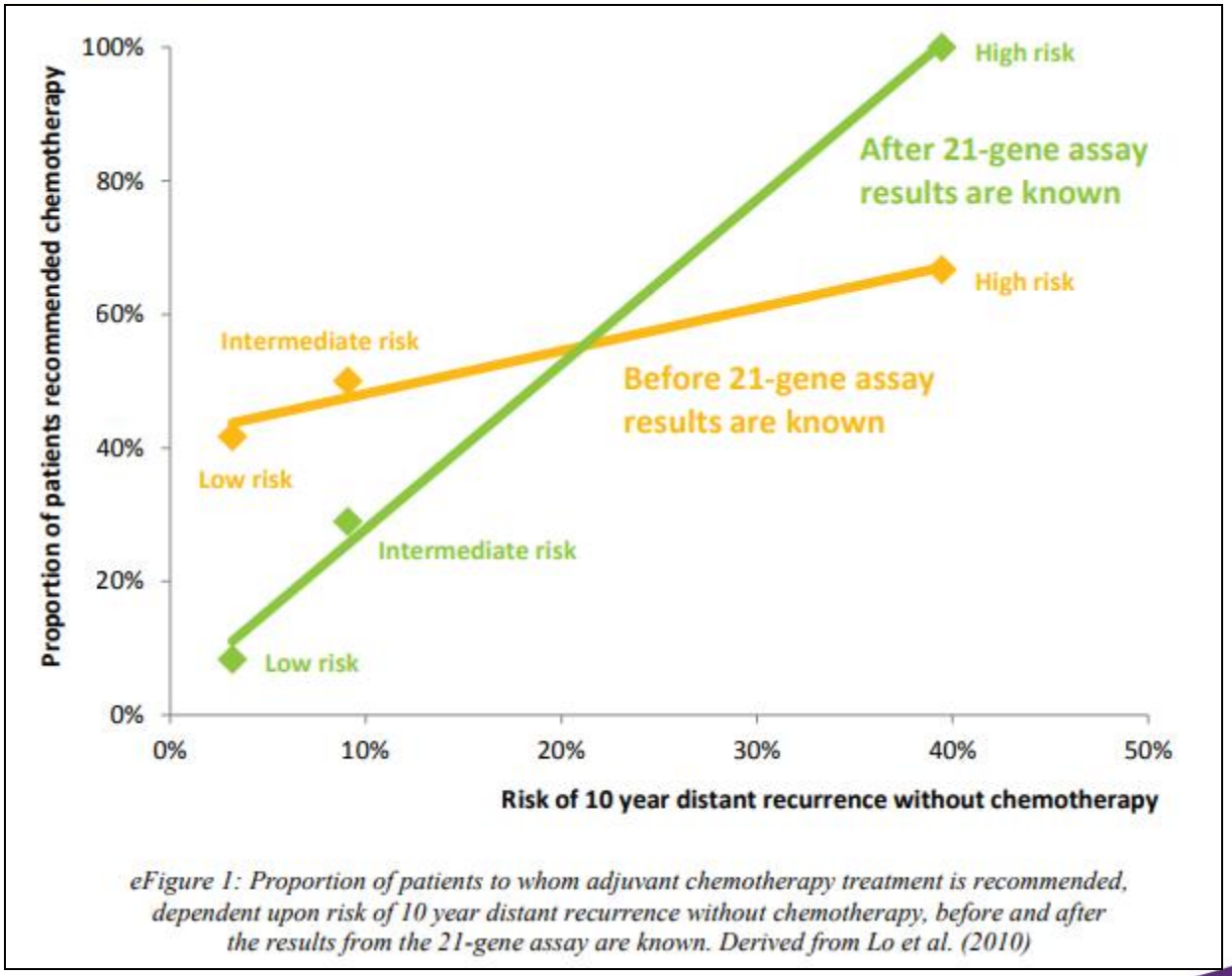
ORIGINAL REPORT

Prospective Multicenter Study of the Impact of the 21-Gene Recurrence Score Assay on Medical Oncologist and Patient Adjuvant Breast Cancer Treatment Selection

Shelly S. Lo, Patricia B. Mumby, John Norton, Karen Rychlik, Jeffrey Smerage, Joseph Kash, Helen K. Chew, Ellen R. Gaynor, Daniel F. Hayes, Andrew Epstein, and Kathy S. Albain

- Enrolled 89 patients
- Collected data pre- and post-test:
  - Medical oncologist adjuvant treatment recommendation and confidence
  - Patient's treatment choice
  - Patient decisional conflict, anxiety, QoL

Subsequently used to inform adjuvant chemotherapy decisions in decision model



eFigure 1: Proportion of patients to whom adjuvant chemotherapy treatment is recommended, dependent upon risk of 10 year distant recurrence without chemotherapy, before and after the results from the 21-gene assay are known. Derived from Lo et al. (2010)

Paulden et al. (2013) Cost-Effectiveness of the 21-Gene Assay for Guiding Adjuvant Chemotherapy Decisions in Early Breast Cancer. Value in Health. 16(5), 729-39.

Key differences between trials and models for evaluating the utility of medical tests.

Randomized trials	Models
Can compare only a few strategies	Can compare many strategies
Evaluates intended and unintended effects	Evaluates modeled effects only
No assumptions	Assumptions necessary
Expensive	Less costly
Time for follow-up needed	Model-building time only

Clin Chem, Volume 58, Issue 12, 1 December 2012, Pages 1636–1643, https://doi.org/10.1373/clinchem.2012.182576

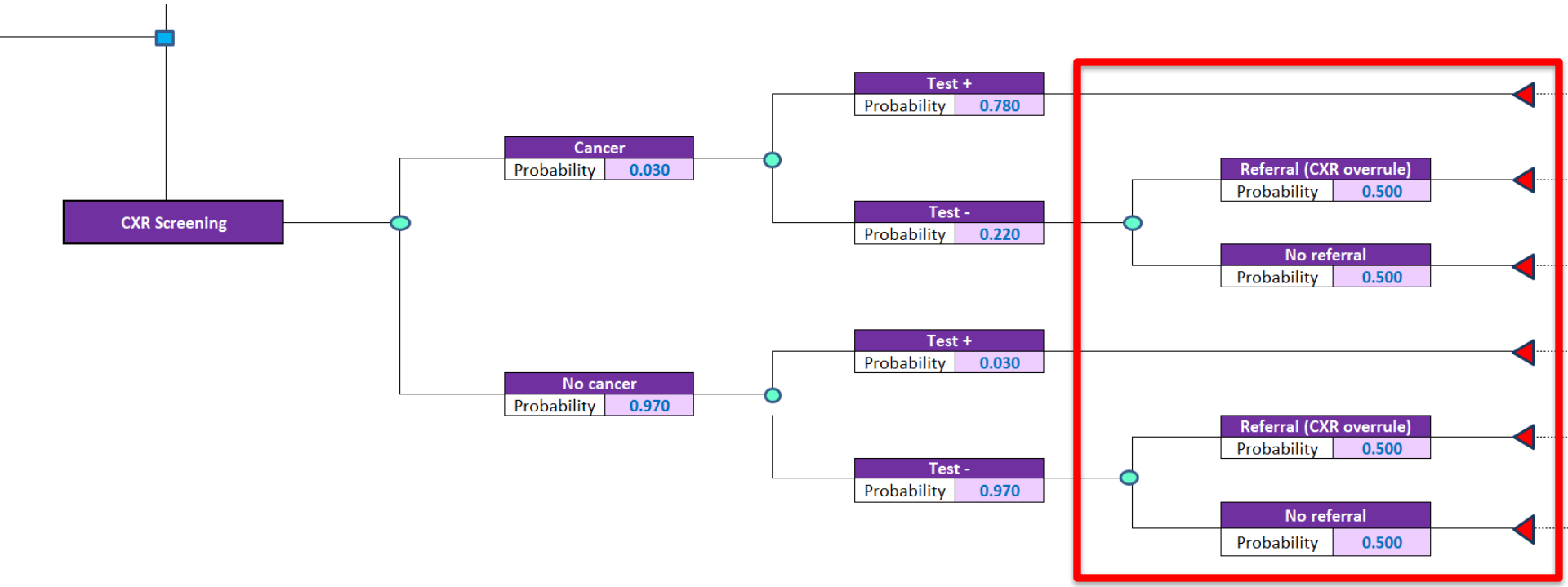
- The assumptions we make when linking accuracy to patient outcomes can significantly influence the cost-effectiveness of the test

- We're going to explore the impact of different modelling assumptions/issues:
  - Adherence to test results
  - Impact of missing lung cancer
  - Changing treatment costs
  - Impact of overdiagnosis

Thinking back to our comparison of CXR to LDCT for lung cancer screening – how could we adapt our model to allow for some negative CXRs to be referred anyway?

Example 3 – Decision Tree

Added in a probability that a negative CXR will be referred anyway



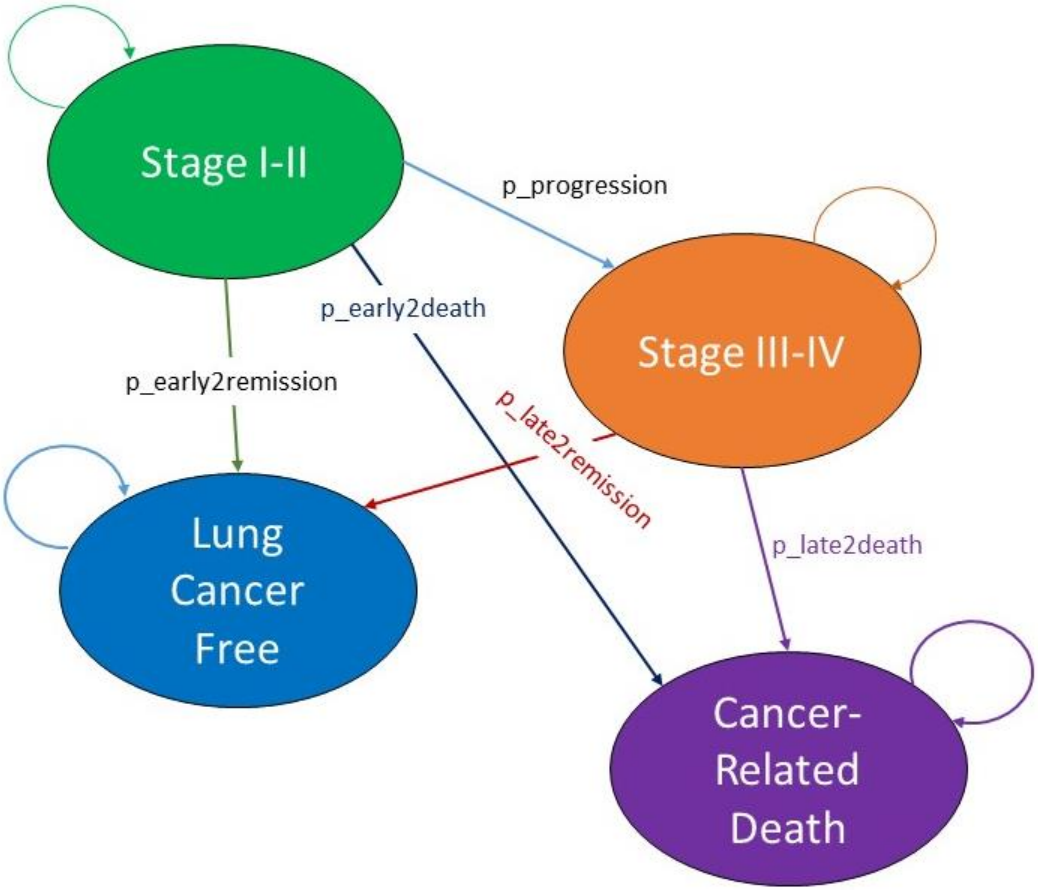
Base Case Results

Strategy	Total cost	Total QALYs	Incremental cost	Incremental QALY	ICER	INHB	INMB
Chest XR	$1,031	4.46	-	-	-	-	-
LDCT	$1,154	4.46	$123	0.0040	$31,090	0.0027	$272

CXR Adherence Scenario Analysis

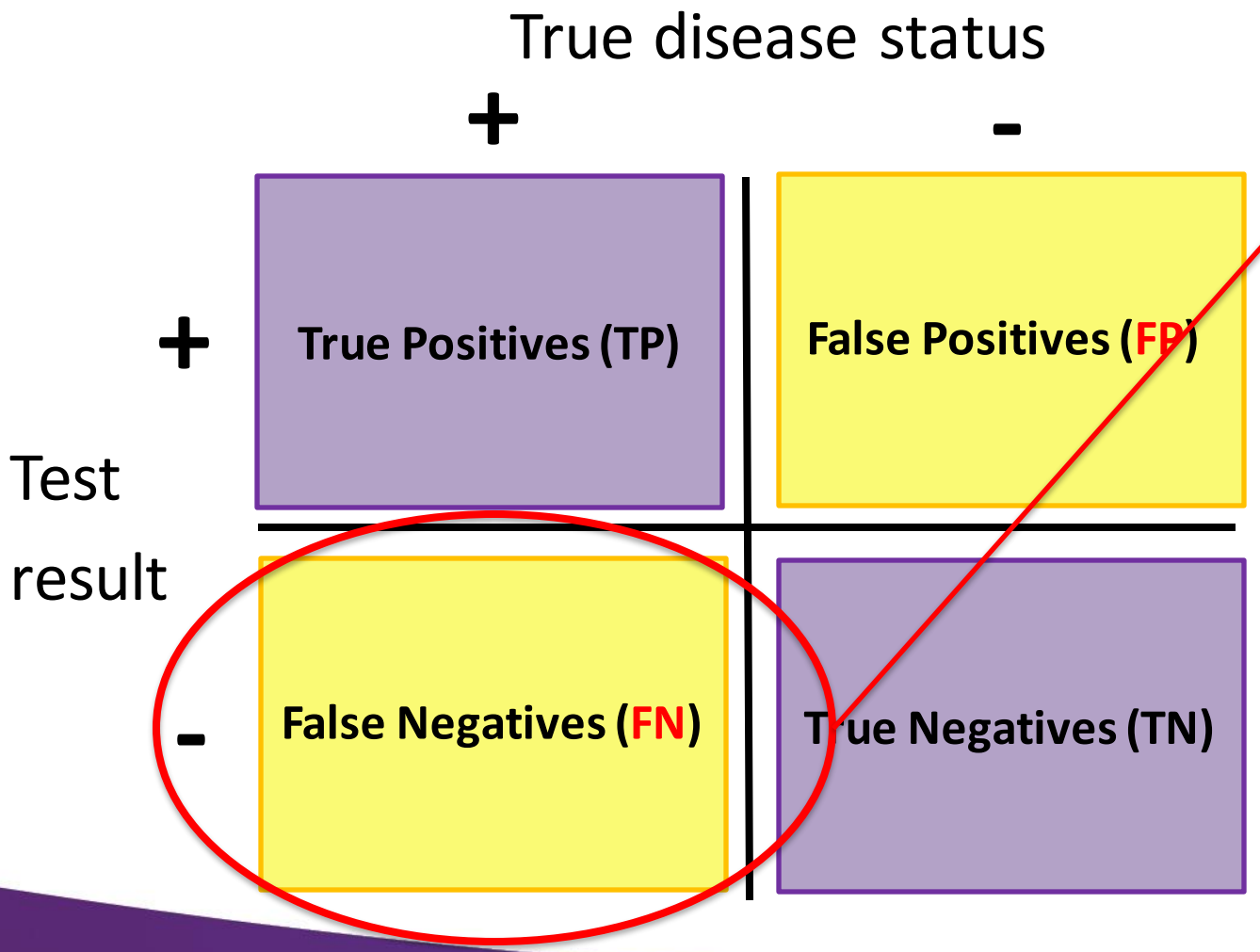
Strategy		Total cost	Total QALYs	Incremental cost	Incremental QALY	ICER	INHB	INMB
CXR test		$1,259	4.45	-	-	-	-	-
LDCT		$1,154	4.46	-$105	0.0074	LDCT dominates	0.0085	$847

- Decision tree modelling is inflexible when capturing long-term costs and outcomes
- Linking a Markov model to the tree instead provides greater flexibility
- Look at the example Markov model in Excel which captures the downstream outcomes for those with lung cancer
- We will use this to explore different scenarios in terms of the downstream impact of tests on costs and outcomes


Stage I-II
p_progression
Stage III-IV
p_early2death
p_early2remission
p_late2remission
p_late2death
Lung Cancer Free
Cancer-Related Death

Base case results

		COSTS	OUTCOMES	NET BENEFIT
CXR		$ 24,877	11.322	$ 1,107,366
LDCT		$ 26,835	13.487	$ 1,321,900
Incremental		$ 1,958		
		for	2.165	QALYs
ICER		$ 904.27	per QALY	



What are the cost and outcome implications of missing a lung cancer?

- CXR misses a lot more cancers than LDCT
- But what if these cancers are generally very small and slow growing?
- Our assumption that all cancers missed by CXR would progress to stage III/IV before being diagnosed would be incorrect
- What do you think would happen if we assumed that only a small proportion of those missed on CXR would progress to late stage cancer?

See Example 1 – Markov Model

Base Case Results

Strategy	Total Cost	Total QALYs	Incremental Cost	Incremental QALYs	ICER	INHB	INMB
CXR	$24,877	11.32	-	-	-	-	-
LDCT	$26,835	13.49	$1,958	2.165	$904.27	98,044	$214,534

False Negative on CXR Scenario Analysis

Strategy	Total Cost	Total QALYs	Incremental Cost	Incremental QALYs	ICER	INHB	INMB
CXR	$26,903	13.73	-	-	-	-	-
LDCT	$26,835	13.49	-$68	-0.247	$276.48	100,068	-$24,595

- Unlike decision trees, Markov modelling allows you to change the costs and utility values associated with a specific health state
- In our Markov model, the data on the cost treating different stages of lung cancer is uncertain
- What do you think would happen if we reduce the cost of treating stage I/II cancer?

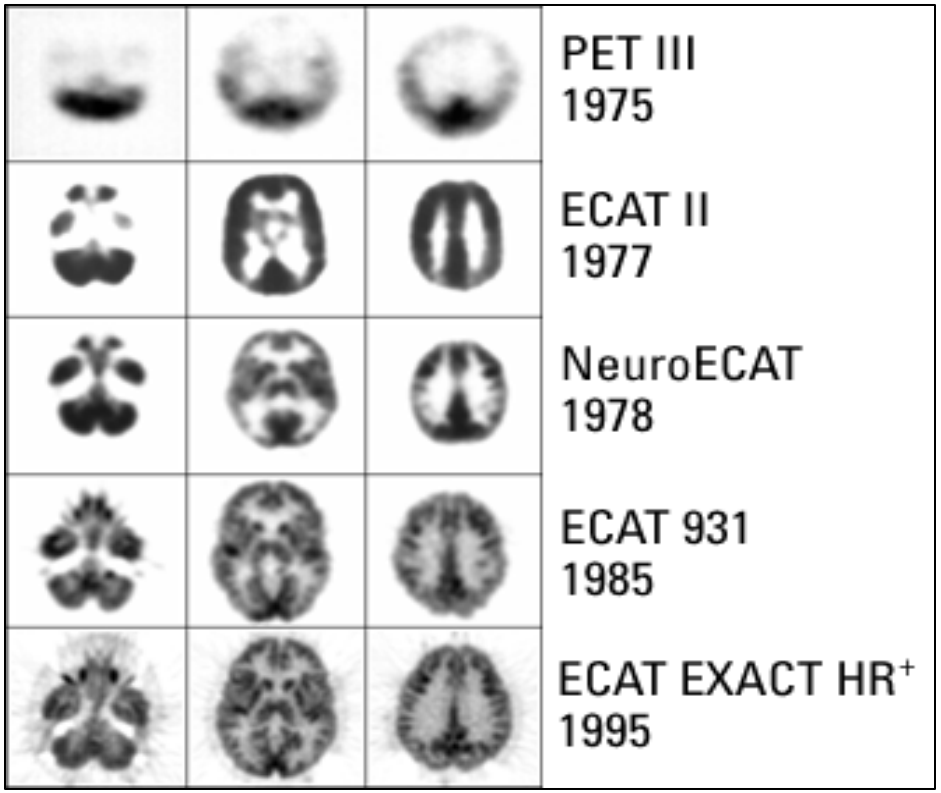
See Example 2 – Markov Model

Base Case Results

Strategy	Total Cost	Total QALYs	Incremental Cost	Incremental QALYs	ICER	INHB	INMB
CXR	$24,877	11.32	-	-	-	-	-
LDCT	$26,835	13.49	$1,958	2.165	$904	98,044	$214,534

Treatment Cost Scenario Analysis

Strategy	Total Cost	Total QALYs	Incremental Cost	Incremental QALYs	ICER	INHB	INMB
CXR	$21,288	11.32	-	-	-	-	-
LDCT	$22,446	13.49	$1,218	2.165	$563	98,784	$215,273



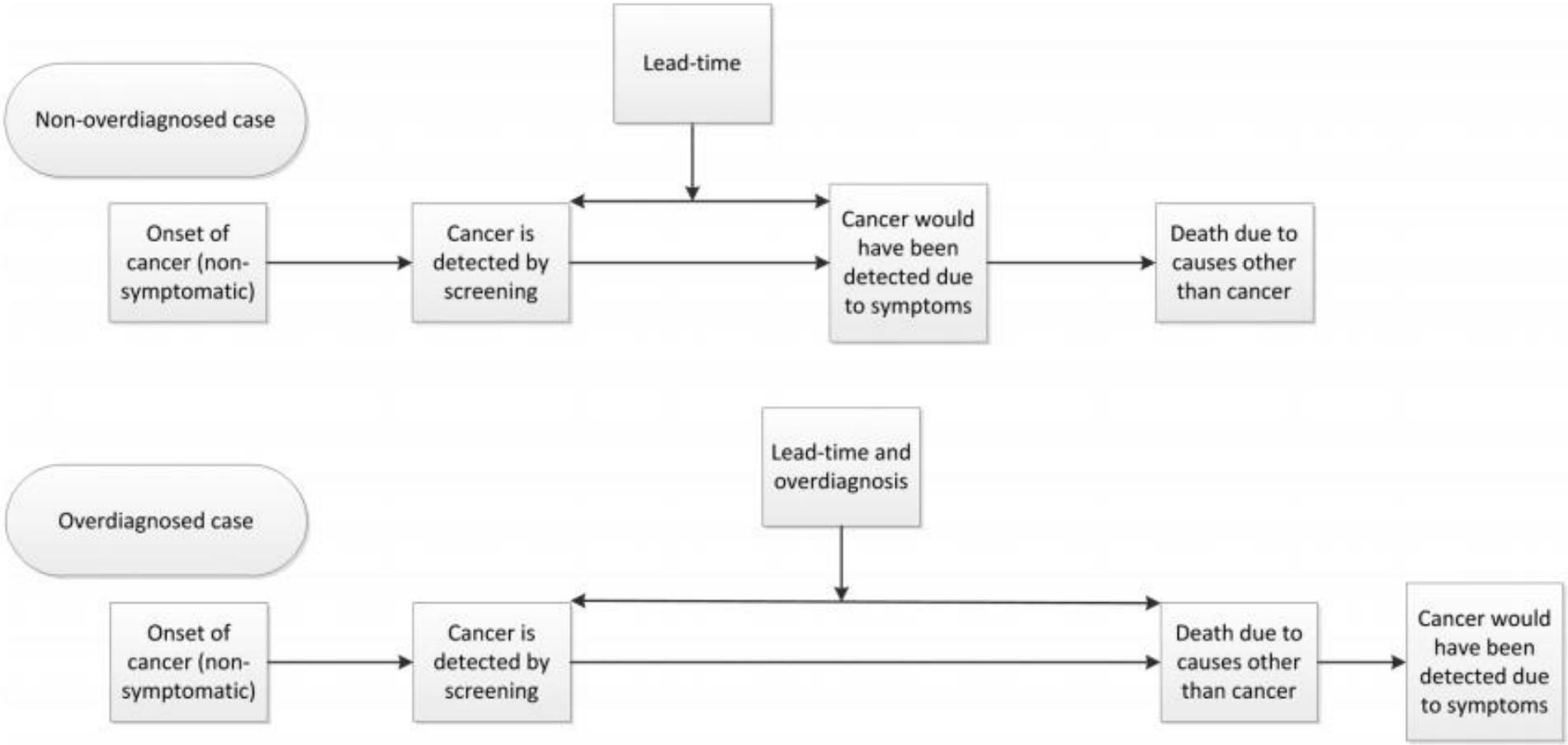
- Sensitivity of imaging tests has vastly improved over the years
- Implicitly widens the definition of disease
- We can now pick up minor abnormalities which may not ever progress to cause symptoms

What is overdiagnosis?

“The diagnosis of disease that will never cause symptoms or harm during a patient’s lifetime”

Potential consequences:

- Unnecessary anxiety
- Harms from excessive treatment
- Increased healthcare costs
- Risk that those who don’t need treatment are prioritised over those who do



ten Haaf K, de Koning HJ. (2015) Overdiagnosis in lung cancer screening: why modelling is essential. J Epidemiol Community Health, 69:1035-1039.

- Smaller nodules show up on LDCT compared to CXR and therefore overdiagnosis is a possibility
- If we only had data on lung cancer-specific mortality, and survival is better for LDCT

See example 3 – Markov model

- To fully account for overdiagnosis, we would need to add a 'not clinically significant lung cancer' health state to our model

Base Case Results

Strategy	Total Cost	Total QALYs	Incremental Cost	Incremental QALYs	ICER	INHB	INMB
CXR	$24,877	11.32	-	-	-	-	-
LDCT	$26,835	13.49	$1,958	2.165	$904.27	98,044	$214,534

Overdiagnosis Scenario Analysis

Strategy	Total Cost	Total QALYs	Incremental Cost	Incremental QALYs	ICER	INHB	INMB
CXR	$24,877	11.32	-	-	-	-	-
LDCT	$27,214	13.65	$2,337	2.322	$1,002	97,666	$230,839

Patz et al. propose that we define overdiagnosis as:

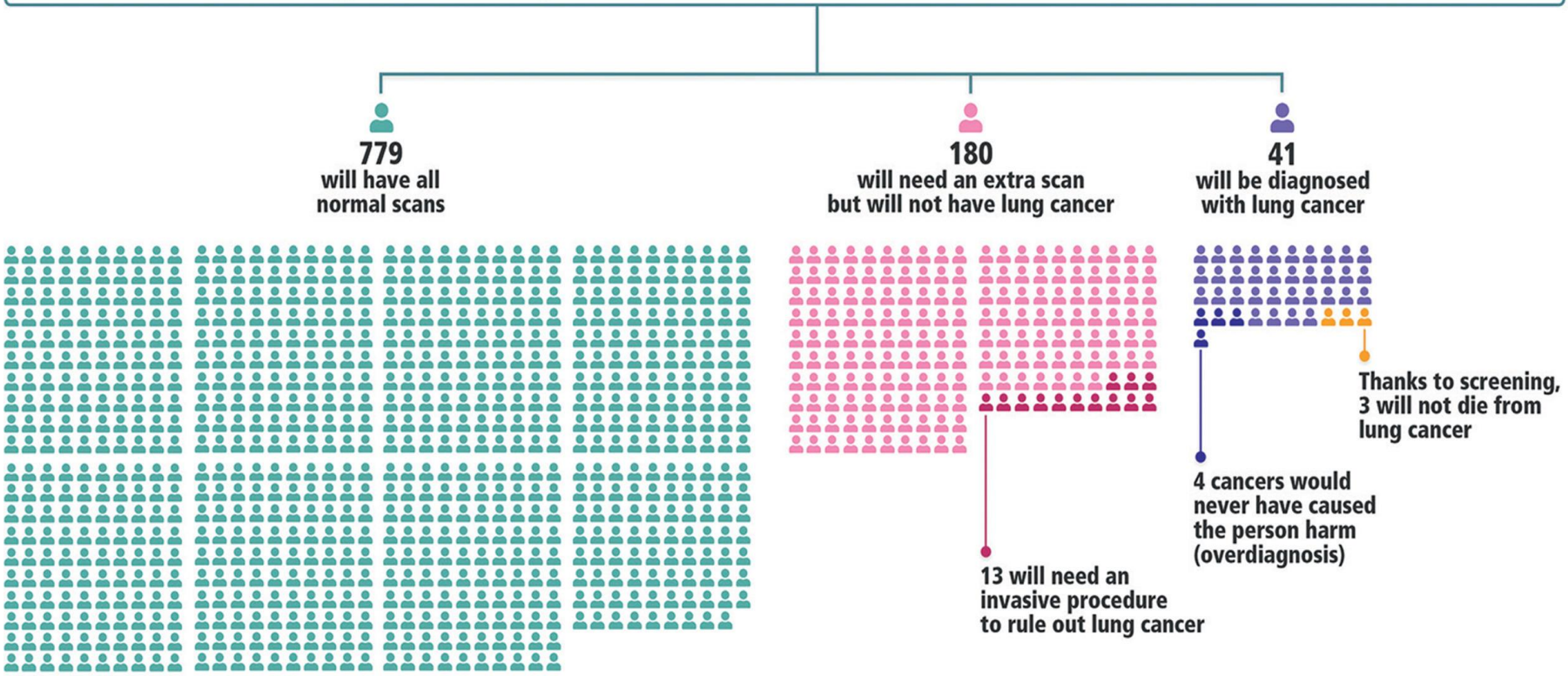
"the excess lung cancers detected by LDCT divided by all lung cancers detected by screening in the LDCT arm"

Requires long follow-up because there is typically a "catch-up" period in the non-screened arm

For NLST, the probability that a LDCT screening-detected lung cancer was an overdiagnosis was 18.5% (95% CI, 5.4%-30.6%)

JAMA Intern Med. 2014;174(2):269-274.
doi:10.1001/jamainternmed.2013.12738

Any screening programme has potential benefits and harms. For lung cancer screening, if 1000 eligible individuals are screened 3 times, it is estimated that:



The Lancet Respiratory Medicine 2019 7655-656DOI: (10.1016/S2213-2600(19)30136-5)

Conclusions

- Decision analytic modelling provides a useful framework for linking diagnostic accuracy data to evidence on downstream costs and patient outcomes
- These models are sensitive to the assumptions made when linking the data
- Capturing the implications of FNs and FPs can be challenging – may involve specific studies
- It's important to capture possible harms (e.g. overdiagnosis)