III. Digital Twin AI for Precision Screening

Professor Tony Hsiu-Hsi Chen

PART 1: Stochastic Process for Personalized Multi-State Disease Natural History.

Amy

Wenly

P

Abbie

Chen-Yang

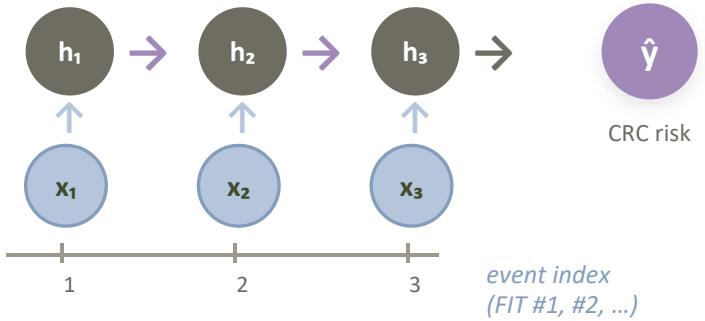
Sam

BRIDGING MACHINE LEARNING & STOCHASTIC PROCESSES

LSTM Meets the Stochastic Process

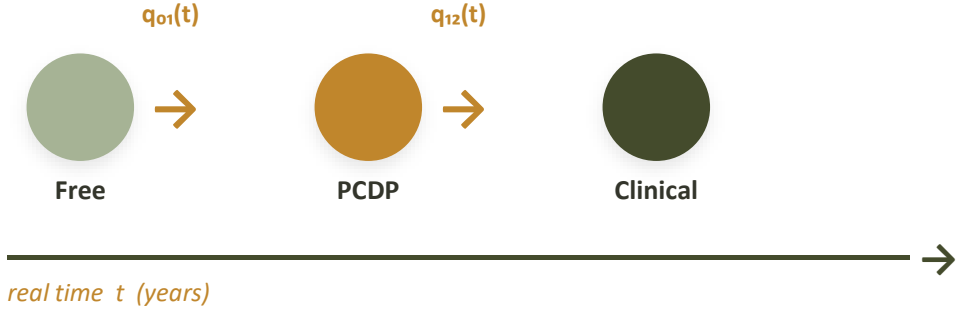
A classifier indexed by observations — versus a natural-history model indexed by time.

LSTM — a Sequential Classifier



- Hidden state h_t is defined by the inputs X.
- "Time" = order in the sequence — no real clock.
- Blind to the gap Δt between screenings.
- Goal: a classifier → ŷ = predicted risk.

Stochastic Process — Real Continuous Time



- Latent state evolves in real, continuous time t.
- Intensities q(t) & sojourn times are true durations.
- The interval between screens carries information.
- Goal: model the natural history → time-resolved risk.

The bridge — the LSTM's hidden process is defined by the data ***X***; the stochastic process is defined by ***time t***. Give the learner a real clock, and the two become one model.

ROADMAP

The Arc of This Talk

A multi-state journey — from the hidden biology of disease to AI-assisted precision screening.

01

Hidden States

Stochastic processes recover the multi-stage transitions hidden behind a screen.

02

Beyond Markov

Non-homogeneous incidence — with transition rates indexed by covariates.

03

Initiator & Promoter

Personalized factors drive a 3-state natural-history model.

04

Tailored Screening

Genetic & epigenetic factors → an individualized strategy.

THE CHALLENGE

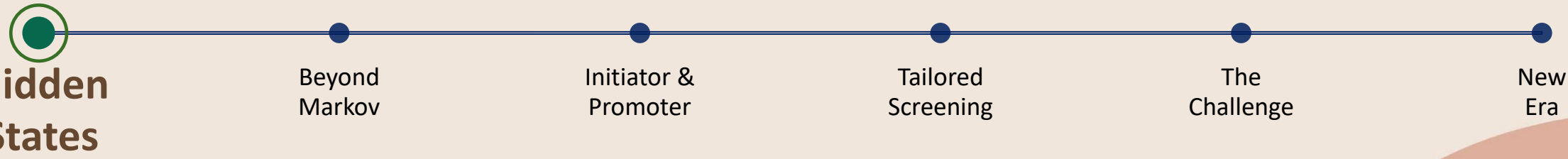
Estimation turns hard — the linear-regression intuition we trust no longer holds.

NEW ERA

① Code-free SAS & Python tools
② Machine learning & artificial intelligence

1. Hidden States

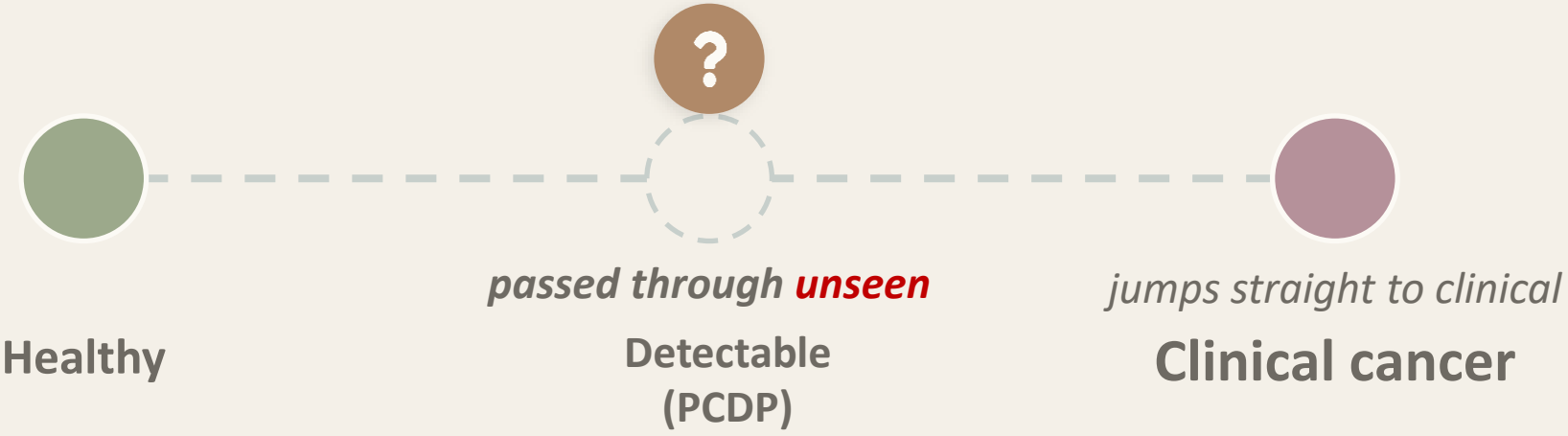
Screening sees only a sliver of the disease process. Stochastic models recover the multi-stage transitions underneath.



What Screening Data Cannot see?

1 Screen-detected, then removed

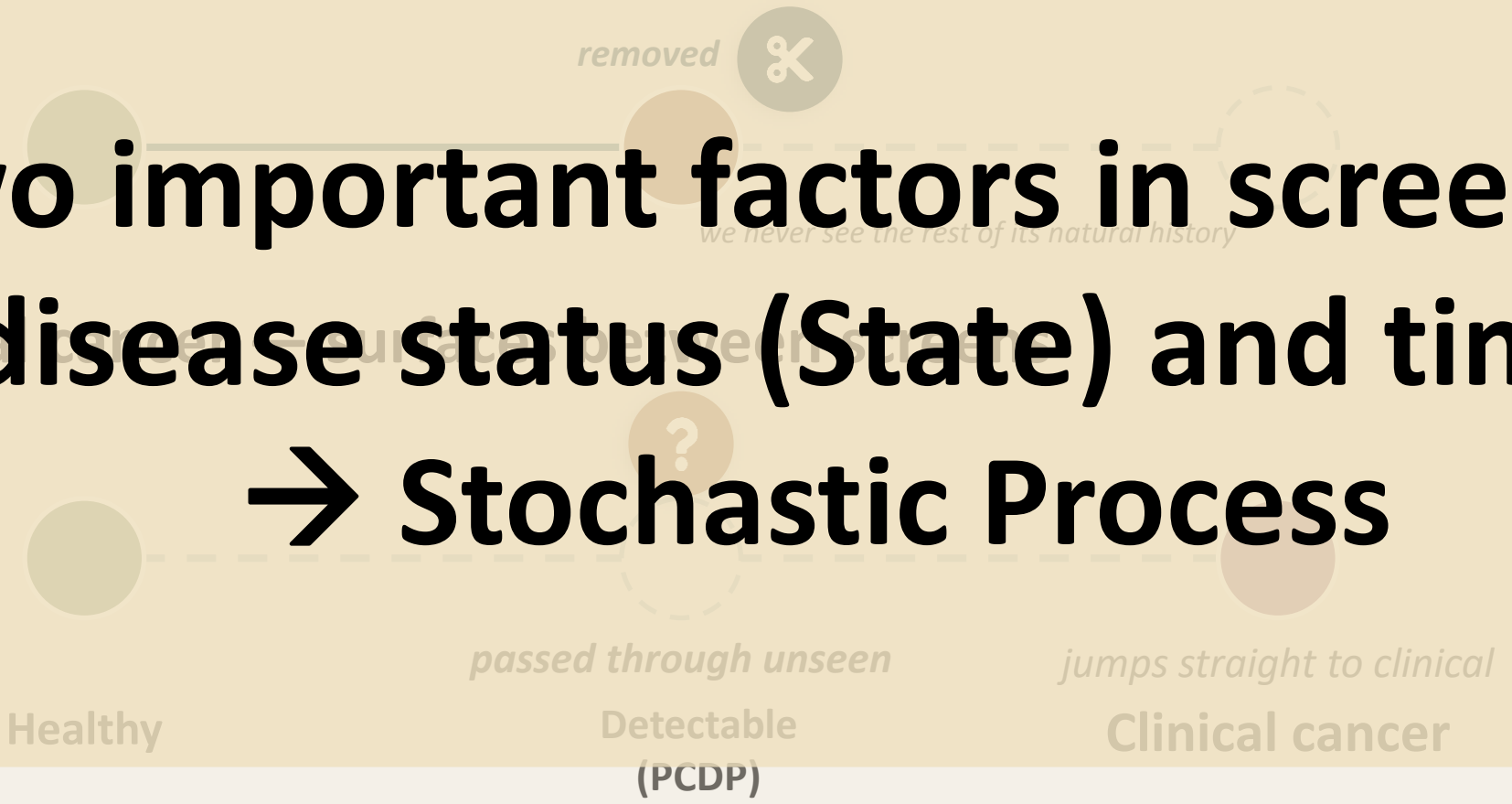
2 Interval cancer — surfaces between screens



What Screening Data Cannot see?

1 Screen-detected, then removed

2 Interval cancers: cases between tests



Two important factors in screening: disease status (State) and time.

→ Stochastic Process

REAL WORLD DATA

TAMCAS Cohort

EFFICACY OF BREAST-CANCER SCREENING FOR FEMALE RELATIVES OF BREAST-CANCER-INDEX CASES: TAIWAN MULTICENTRE CANCER SCREENING (TAMCAS)

Mei-Shu LAI[1], Ming-Fang YEN[2], Hsu-Sung KUO[2], Shin-Lan KOONG[1], Tony Hsiu-Hsi CHEN[3]* and Stephen W. DUFFY[4]

Study Design

Index Cases

Breast Cancer Patients from Hospitals

↓

Period: 1995 to 1998

2nd degree female relatives, aged 35+ (N= **4,865**)

↓

Invited for annual screening

Physical Examination + Mammography + Ultrasound

↙

86 screen-detected cases and 7 interval cancers

Lai MS et al. Int J C; 1998 (78) 21-26

REAL WORLD DATA

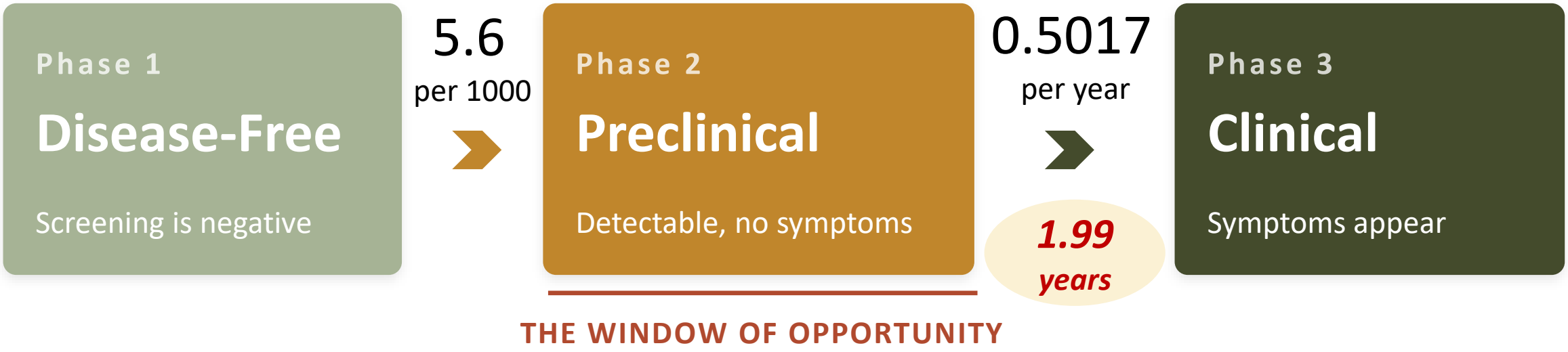
Data on detection mode (state) and time

ID	Screening round	Screening finding	Age	Time
1	1	N	35	-
2	1	N	42	-
3	1	BC	56	-
.	.	.	.	
.	.	.	.	
.	.	.	.	
2629	1	N	52	-
2	2	N	-	2.0
3	2	SD	-	1.8
5	2	N	-	1.5
.	.	.		.
.	.	.		.
.	.	.		.
2400	2	N	-	1.5

Detection mode	Transition history	Number
	1. Prevalent Screen	
Negative cases	$(0 \rightarrow 0, v_m, x_m = 0)$	2598
Positive cases	$(0 \rightarrow 1, v_m, x_m = 1)$	31
	1. Second Screen	
Negative cases	$(0 \rightarrow 0, t_{2i} - t_{1i}, y_{2i} = 0)$	572
Positive cases	$(0 \rightarrow 1, t_{2i} - t_{1i}, y_{2i} = 1)$	3

WINDOW OF OPPORTUNITY

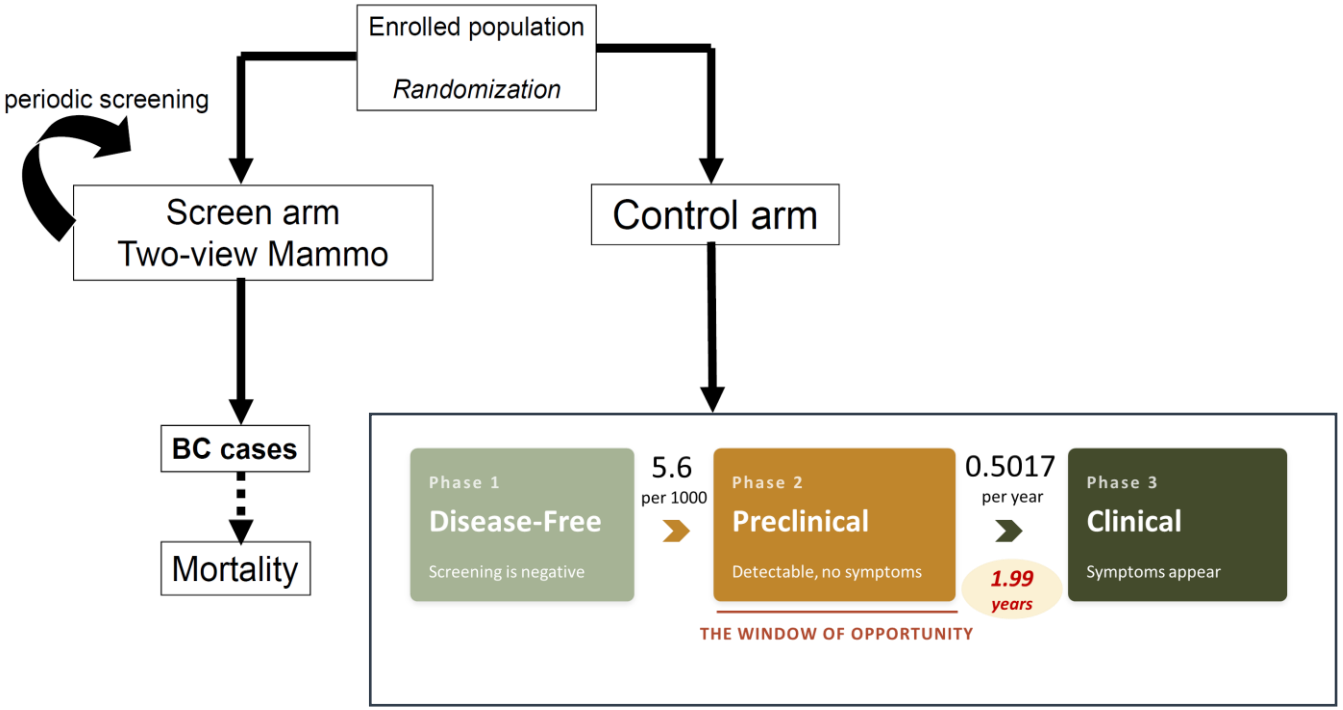
Where Screening Can Act



Sojourn time = *how long disease stays detectable but silent in Phase 2*

Application

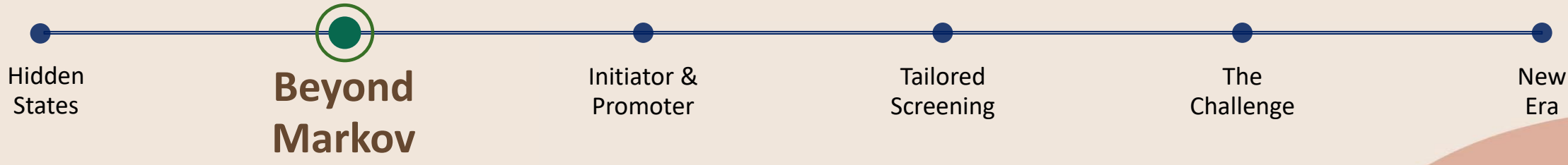
Stochastic Twin Generator



Screening regime	Year	Prevalent cases	Incident screen-detected cases (A)	Interval cancers (B)	$\frac{(B)}{(A)+(B)} \times 100\%$
Annual	0	28.5			22.8
	1		11.5	3.3	
	2		11.4	3.3	
	3		11.3	3.2	
	4		11.2	3.2	
	5		11.2	3.2	
	6		11.1	3.3	
Total		28.5	67.7	19.5	
Two-yearly	0	28.5			38.3
	2		18.2	11.2	
	4		17.9	11.1	
	6		17.6	10.9	
Total		28.5	53.7	33.2	
Three-yearly	0	28.5			50.0
	3		22.0	21.9	
	6		21.4	21.4	
Total		28.5	43.4	43.3	
Control		27.6	—	88.4	

2. Beyond the Markov Assumption

Incidence is not homogeneous over time/ across individuals — so we let transition rates vary, and index them by time/covariates.



WHY CONSTANT RATES FAIL

Homogeneous vs Non-Homogeneous

	Homogeneous	Non-Homogeneous
Transition rates	Constant over time	**Change with age & history**
Risk profile	One-size-fits-all	**Individualized**
Clinical realism	Low	**High**

STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:3369–3382 (DOI: 10.1002/sim.1277)

Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan

Hsin-Ju Hsieh[1], Tony Hsiu-Hsi Chen[2,*,†] and Shu-Hui Chang[1]

[1]*Biostatistics Unit, Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan*
[2]*Institute of Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan*

Real disease accelerates — a flat rate hides who is truly at risk.

EMPIRICAL PROOF

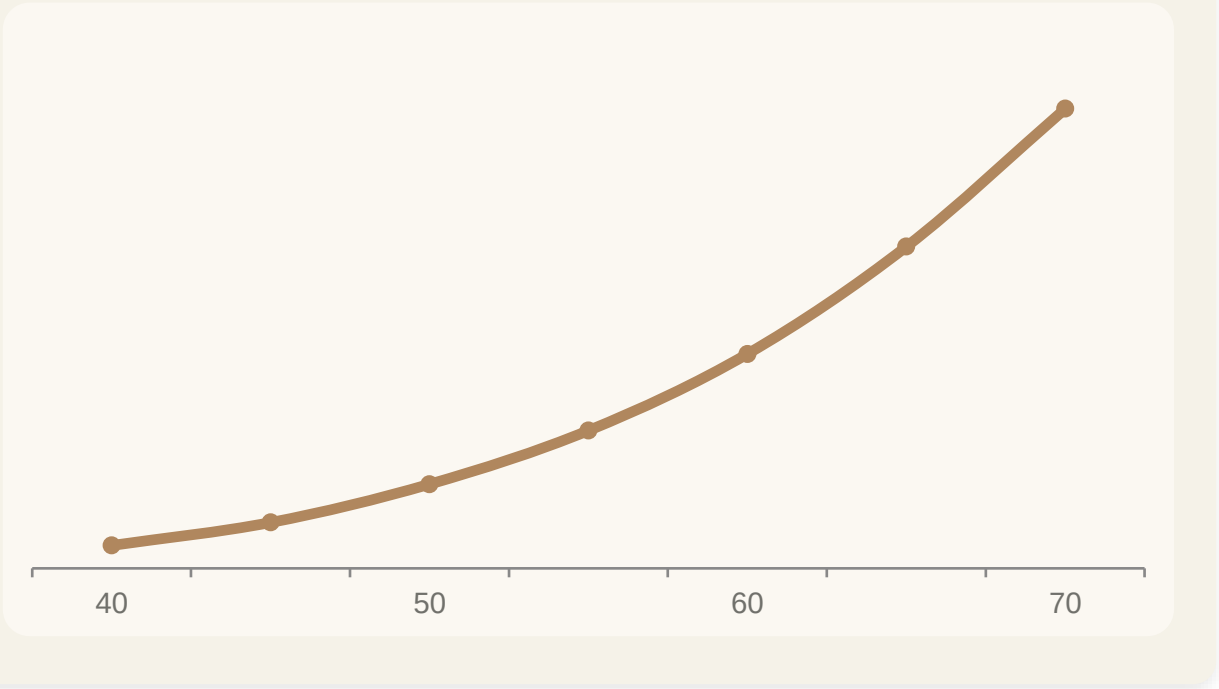
Risk Truly Accelerates

Weibull shape parameter

k = 2.47

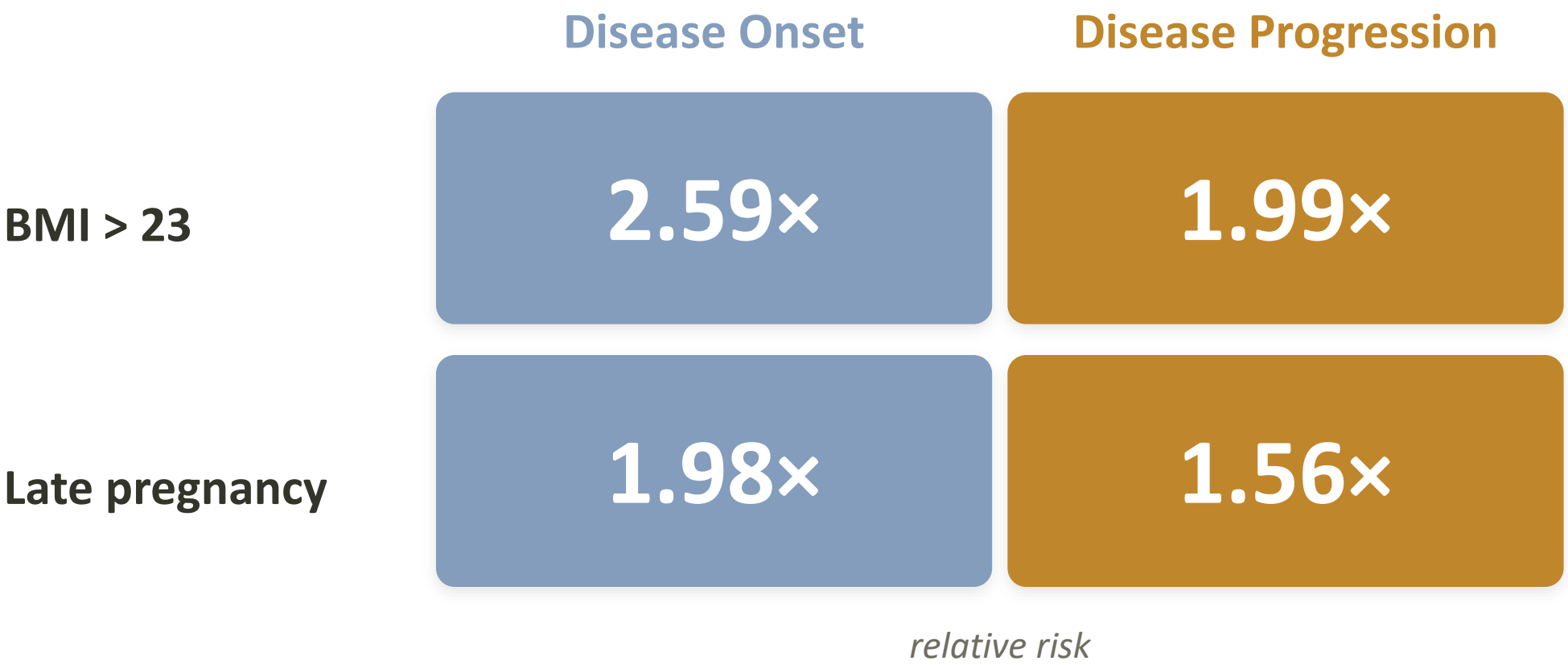
SE = 0.11

k > 1 confirms preclinical incidence rises steeply with age.



MAPPING THE HAZARD RATIOS

Onset and Progression Risk



Obesity and late age act as both initiators and promoters.

Journal of Epidemiology 27 (2017) 98–106

Contents lists available at ScienceDirect

Journal of Epidemiology

journal homepage: http://www.journals.elsevier.com/journal-of-epidemiology/

Original Article

Initiators and promoters for the occurrence of screen-detected breast cancer and the progression to clinically-detected interval breast cancer

Amy Ming-Fang Yen [a, 1], Wendy Yi-Ying Wu [b, c, 1], Laszlo Tabar [d], Stephen W. Duffy [e], Robert A. Smith [f], Hsiu-Hsi Chen [b, *]

3. Initiators & Promoters

Personalized factors enter the 3-state model: who develops disease (initiators), and how fast it advances (promoters).

Hidden States — Beyond Markov — **Initiator & Promoter** — Tailored Screening — The Challeng

COHORT ARCHITECTURE

50,666 Women, Two Eras

Era 1 · 1977–1992

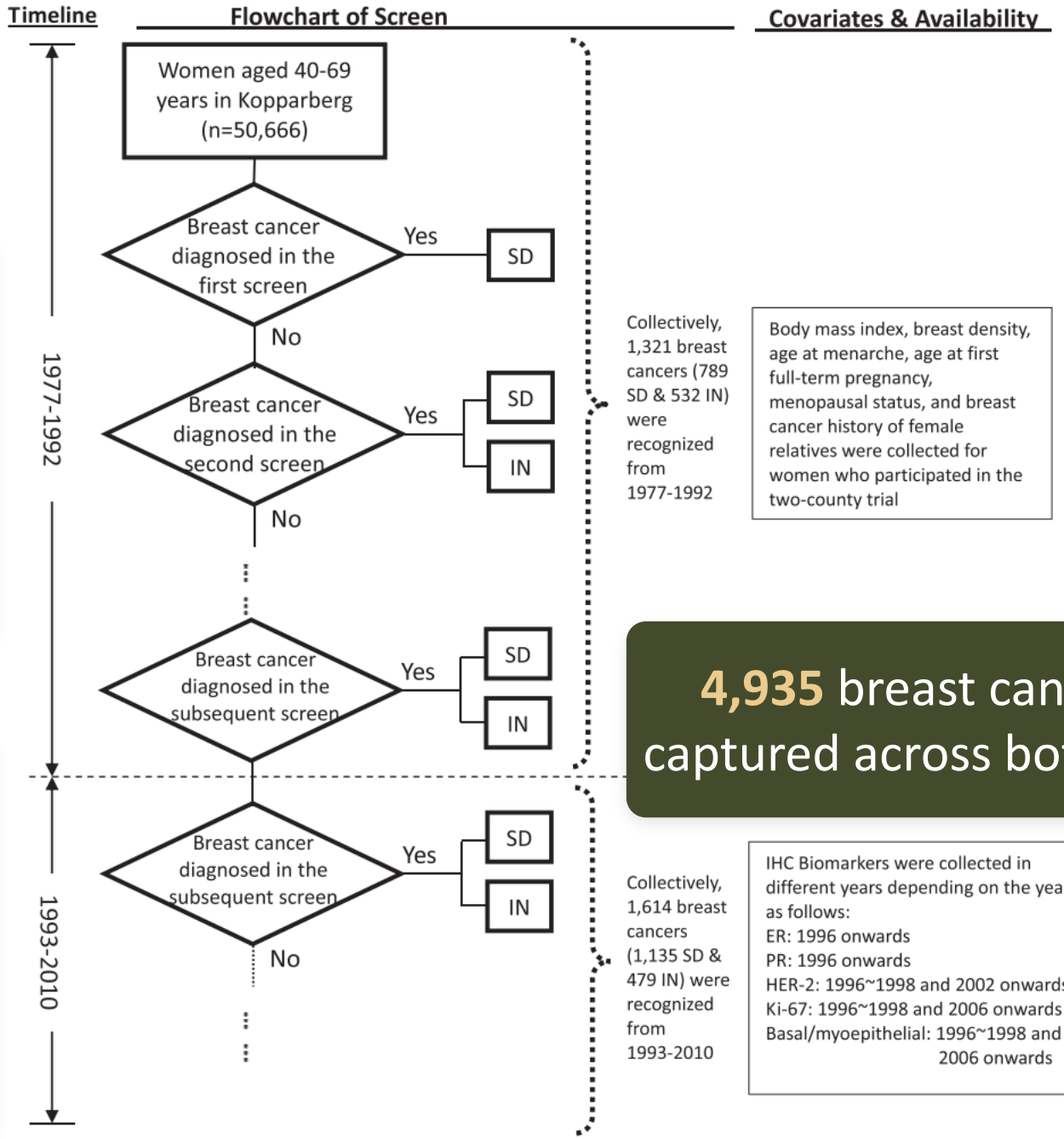
The Trial Period

Active screening every 24 months; conventional risk factors and breast density recorded.

Era 2 · 1999–2010

The Registry Period

Cancer registry linkage adds genes, tumour phenotypes and molecular subtypes.



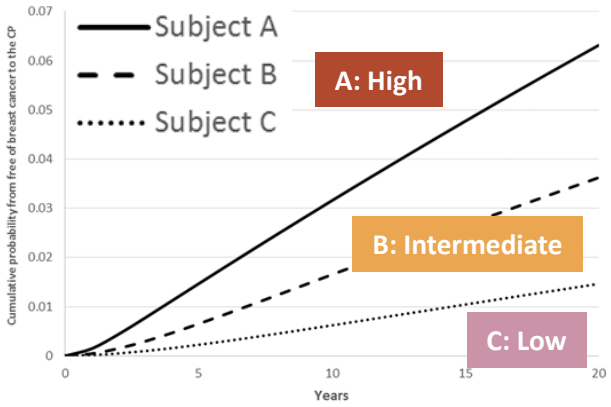
4,935 breast cancers captured across both eras

RISK STRATIFICATION

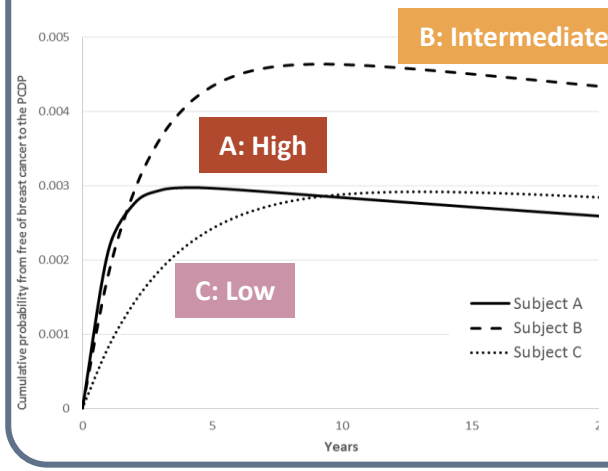
From Profile to Strategy

Profile	BMI	Late preg.	Dense	Family hx	Subtype	Risk
A	>25	Yes	Yes	Yes	Basal-like	High
B	<25	Yes	Yes	No	HER-2+	Intermediate
C	<25	No	No	No	Luminal A	Low

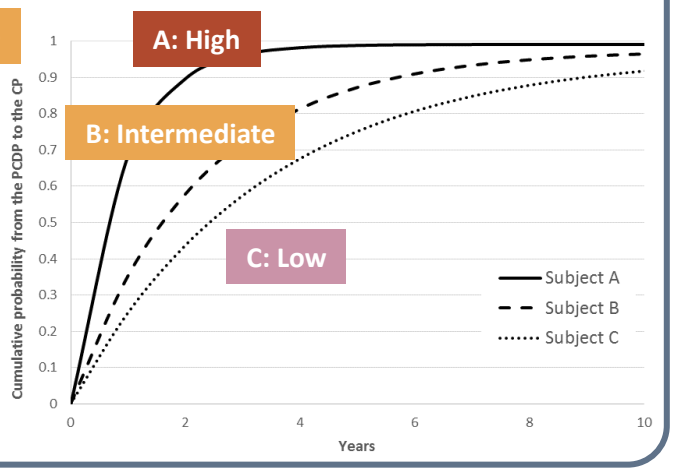
Free of BC → CP



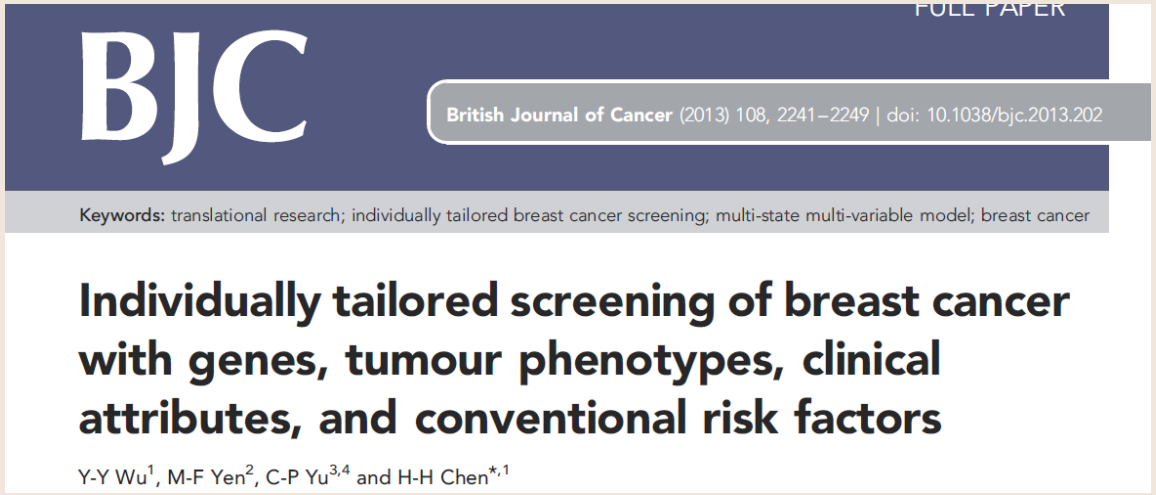
Free of BC → PCDP



PCDP → CP



4. Tailored Screening



FULL PAPER

BJC

British Journal of Cancer (2013) 108, 2241–2249 | doi: 10.1038/bjc.2013.202

Keywords: translational research; individually tailored breast cancer screening; multi-state multi-variable model; breast cancer

Individually tailored screening of breast cancer with genes, tumour phenotypes, clinical attributes, and conventional risk factors

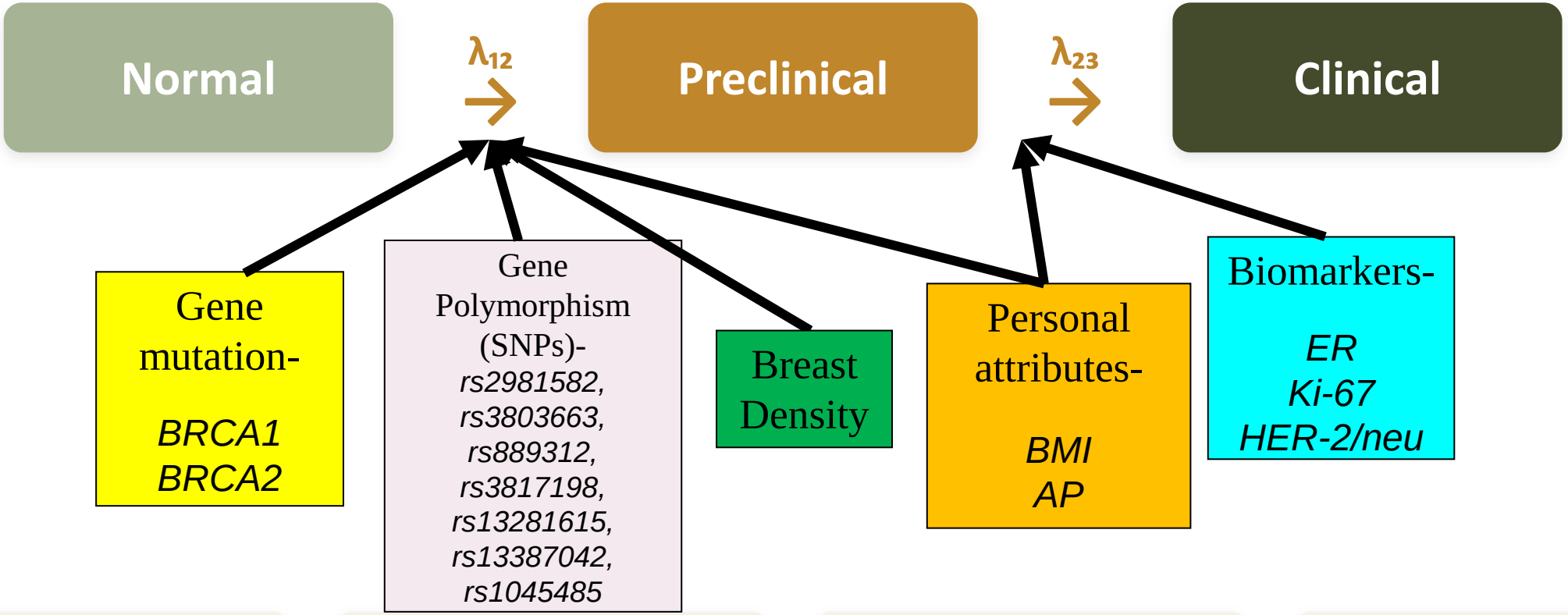
Y-Y Wu[1], M-F Yen[2], C-P Yu[3,4] and H-H Chen[*,1]

Layering in genetic and epigenetic factors turns the model into an individualized screening strategy.

Hidden States → Beyond Markov → Initiator & Promoter → **Tailored Screening** → The Challenge → New Era

LAYERED DETERMINANTS

What Indexes the Rates



High-penetrance genes

Polygenic SNPs

Personal traits

Biomarkers & density

TAILORED PROTOCOL

Screening Matched to Risk

Risk percentile	Start age	Interval	Imaging tool
90–100	30	4.8 months	MRI + Mammography
70–90	35–40	1–1.5 yr	Mammography + Ultrasound
50–70	45–50	2–3 yr	Mammography + Ultrasound
< 50	55–60	4–10 yr	Mammography

Higher risk → *earlier start, tighter interval, more sensitive imaging*

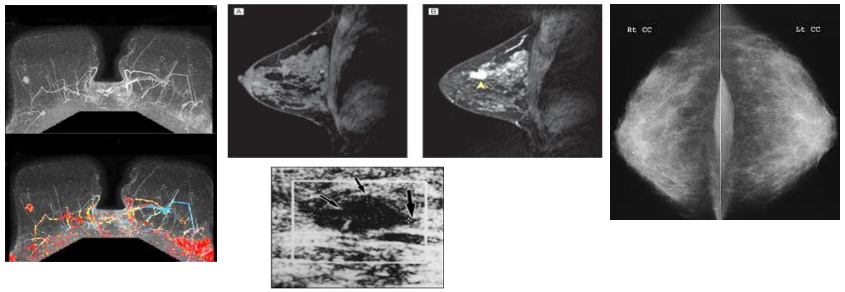
THE PAYOFF

One model — a personalized screening plan for every woman.

Individually Tailored Screening

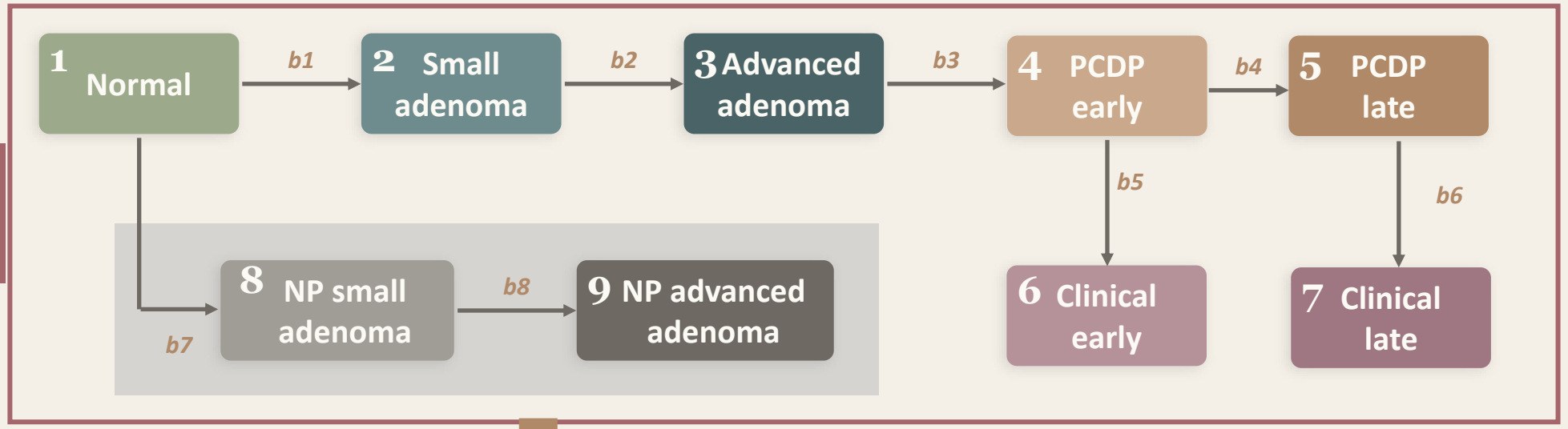
High risk ***Personalized risk for breast cancer*** **Low risk**

Age to being screening	*<30*	*35*	*40*	*50*	*55*	*65*	*65*
Inter-screening interval (year)	*0.5*	*1*	*1.5*	*2*	*3*	*4*	*>4*



Multi-State Disease Model

Multistate Model



Personal features

FIT (f-Hb)	Betel · Alcohol
Age · Sex	Diet
Smoking	Comorbidity

COMBINE ?

Each person's own trajectory through time

High-risk person — fast

Low-risk person — slow

same states, different speed

5. The Challenge

As factors accumulate, estimation grows hard — and the familiar linear-regression relationship no longer holds.

The limits of linear regression

Hidden States — Beyond Markov — Initiator & Promoter — Tailored Screening — **The Challenge** — New Era

6. The New Era

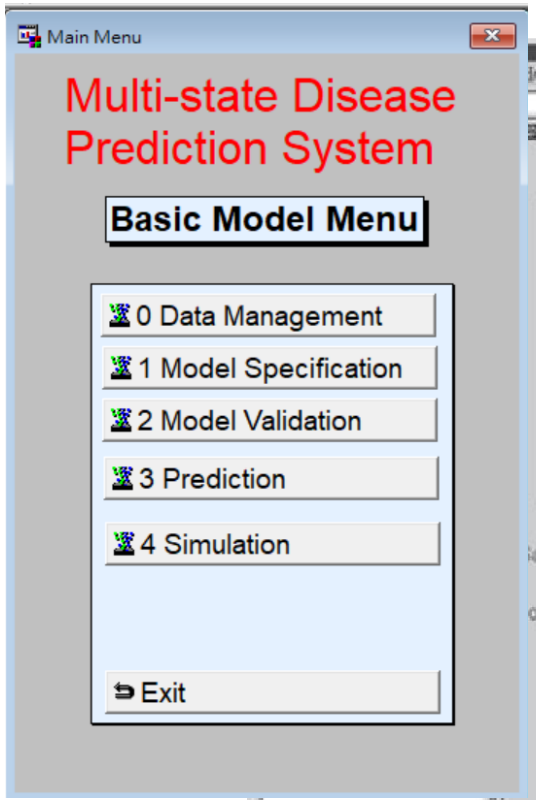
Two answers make the method usable at scale — and push it forward

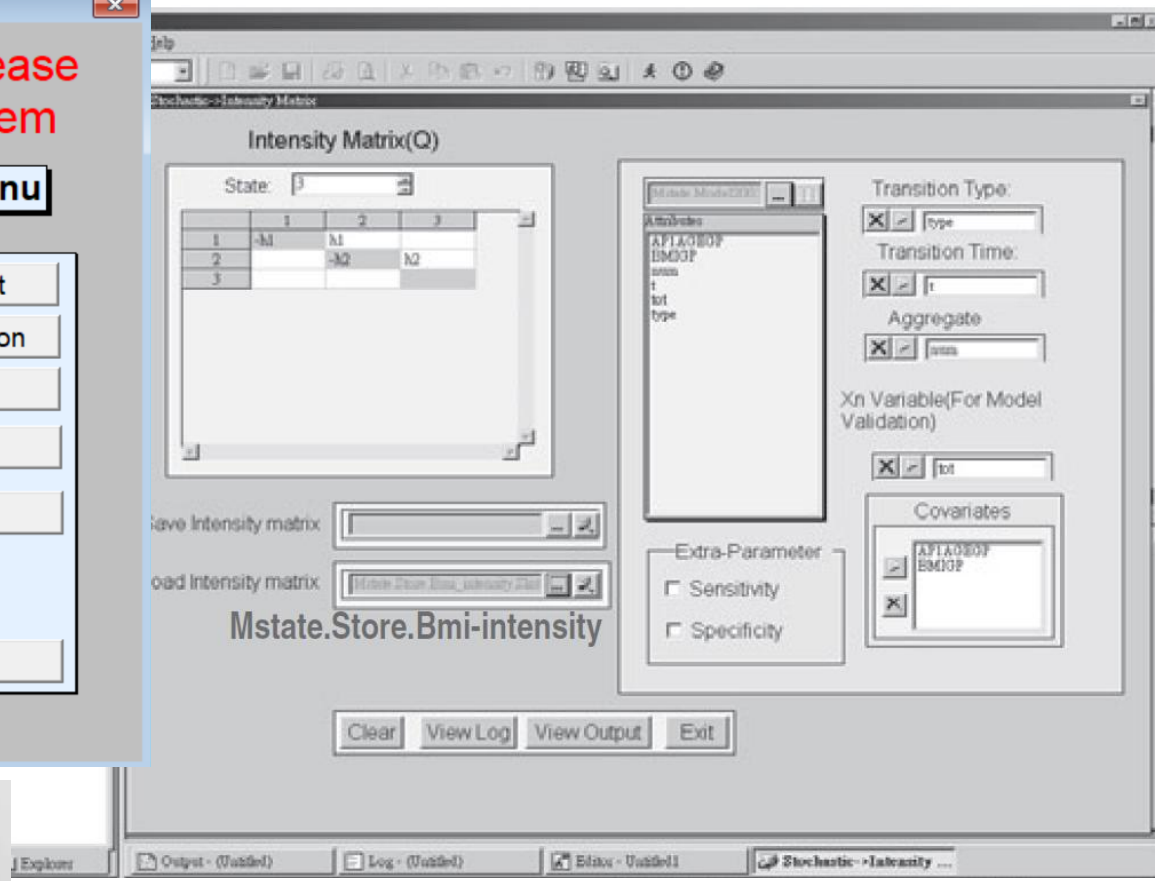
① **Code-free tool SAS / Python**

② **Machine Learning & AI**

Hidden States → Beyond Markov → Initiator & Promoter → Tailored Screening → The Challenge → **New Era**

Multi-state Disease Prediction System





Demonstration:
SAS-based estimation tool

6th Pre-conference workshop in Kosa Hotel, Khon Kaen, Thailand, 2019

M-state demonstration and Python Tool Practice

Chiu-Wen Su
蘇秋文博士
National Taiwan University
臺灣大學

Donlagon Jumparway
National Taiwan University

Estimation and prediction system for multi-state disease process: application to analysis of organized screening regime

Chi-Ming Chang MSc,[1] Wen-Chou Lin MSc,[2] Hsu-Sung Kuo PhD,[3] Ming-Fang Yen MSc[4] and Tony Hsiu-Hsi Chen PhD[5]

Journal of Evaluation in Clinical Practice **13** (2007) 867–881

THE PAYOFF

Multi-state Disease Prediction System

Results | Prediction | Markov Cohort Simulation

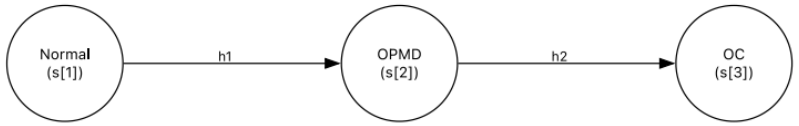
Estimation Results

After convergence, the results page will display:

Status: Converged successfully

2052 sec = 34 minutes

State Transition Diagram



Model Fit Statistics

Sample size: 6308543 observations
Neg Log-Likelihood: 510400.9644
AIC: 1020823.9288
BIC: 1020974.1604

Estimated

with standard errors and 95% confidence intervals.

Transition Rate	Estimate	SE	95% CI
h1	0.0005407	0.000006	[0.0005, 0.0006]
h2	0.0620999	0.000670	[0.0608, 0.0634]

Covariate Effects

Hazard Ratio (HR)

Transition	Covariate	β	SE	HR (e^{β})	95% CI (HR)
h1	mi1_1	0.447761	0.006650	1.5648	[1.5445, 1.5853]

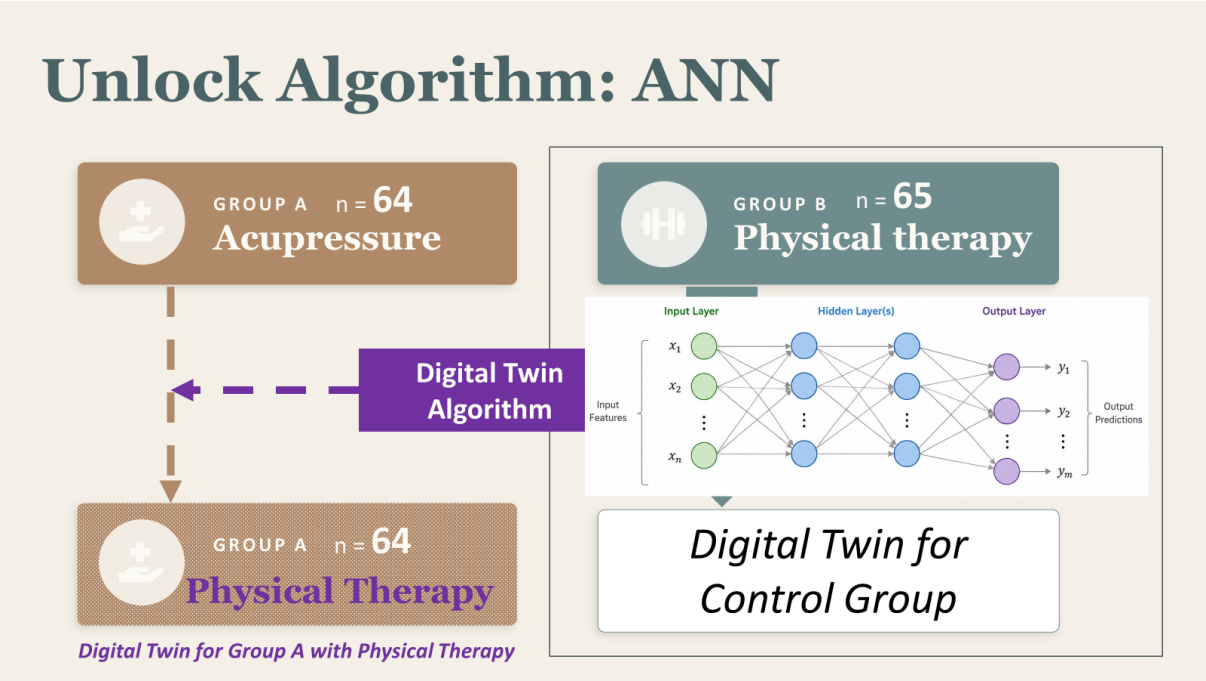
Hand-on Practice:

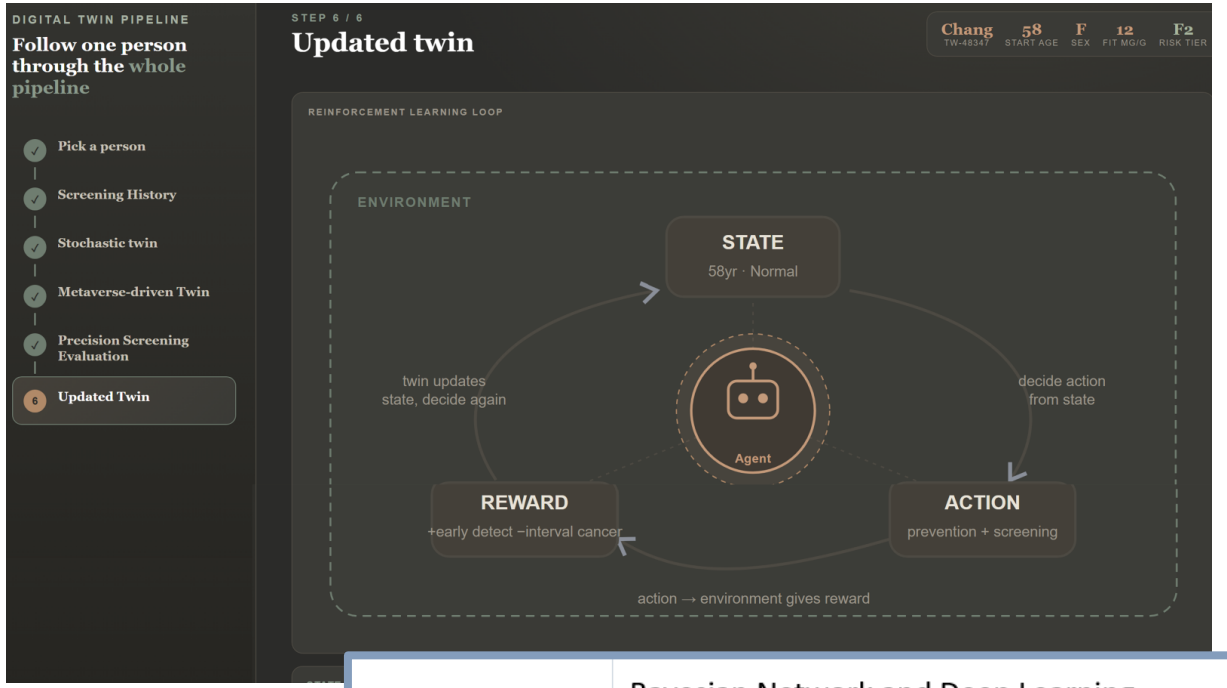
Python-based estimation tool

M-state demonstration and Python Tool Practice

Chiu-Wen Su
蘇秋文博士
National Taiwan University
臺灣大學

Donlagon Jumparway
National Taiwan University

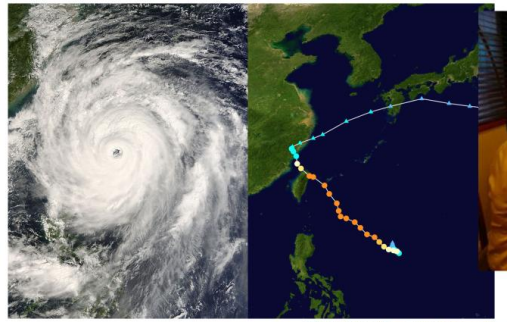




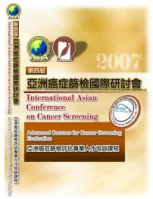
14:10-14:30	Bayesian Network and Deep Learning Approaches for Personalized Multi-State Disease Natural History Modeling
14:30-15:00	AI-Empowered Digital Twin Frameworks for Precision Screening
15:20-15:50	Reinforcement Learning–Enabled Digital Twins for Risk-Adaptive Screening Strategies with LLM
15:50-16:10	Digital Twin AI Demonstration for Precision Screening Applications

Digital Transformation of Cancer Screening

2nd Pre-conference workshop in Taiwan, 2007



Strong level-
Typhoon Krosa (柯莎)



Welcome Hotel
家美飯店

Universities

Clinicians

Community

Taiwan
Cancer Screening Program

NGO

Government