BIG DATA in Health Sciences

Innovations for health care, medical product development and regulatory science

Efficient Sampling-design in Multistate Disease Process for Screening and Surveillance

Dr. Chen-Yang Hsu, Dr. Wen-Feng Hsu

Professor Hsiu-Hsi Chen

Division of Biostatistics, College of Public Health,

National Taiwan University

國立臺灣大學 公共衛生學院

College of Public Health National Taiwan University

Multistate Model and Disease progression

1. State-specific effect of multistate disease progression

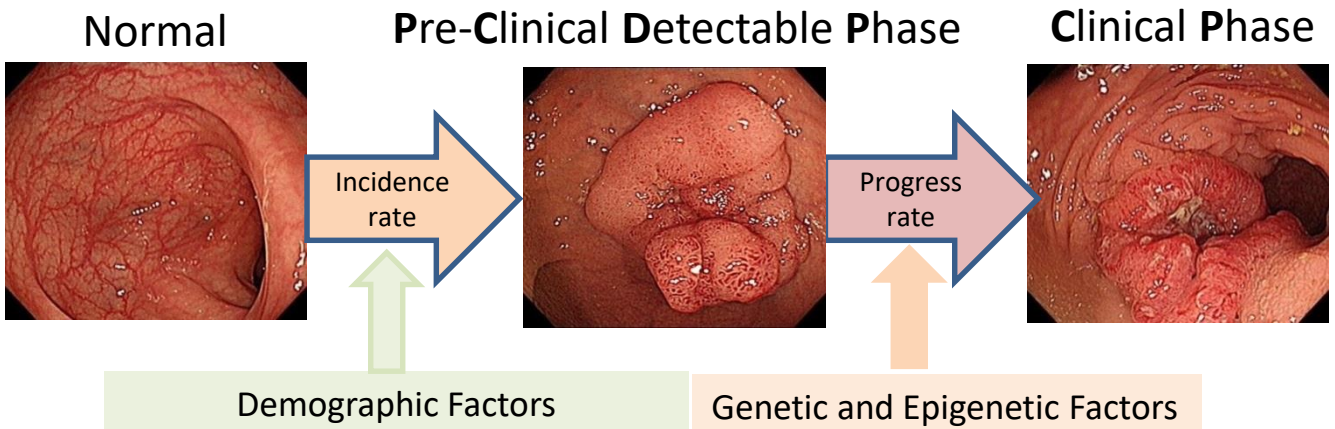
2. State-specific effect of multistate disease progression

3. Risk- and State-based cancer prevention

Efficacy=1-O/E

O: observed rate of disease progression

E: expected rate of disease progression



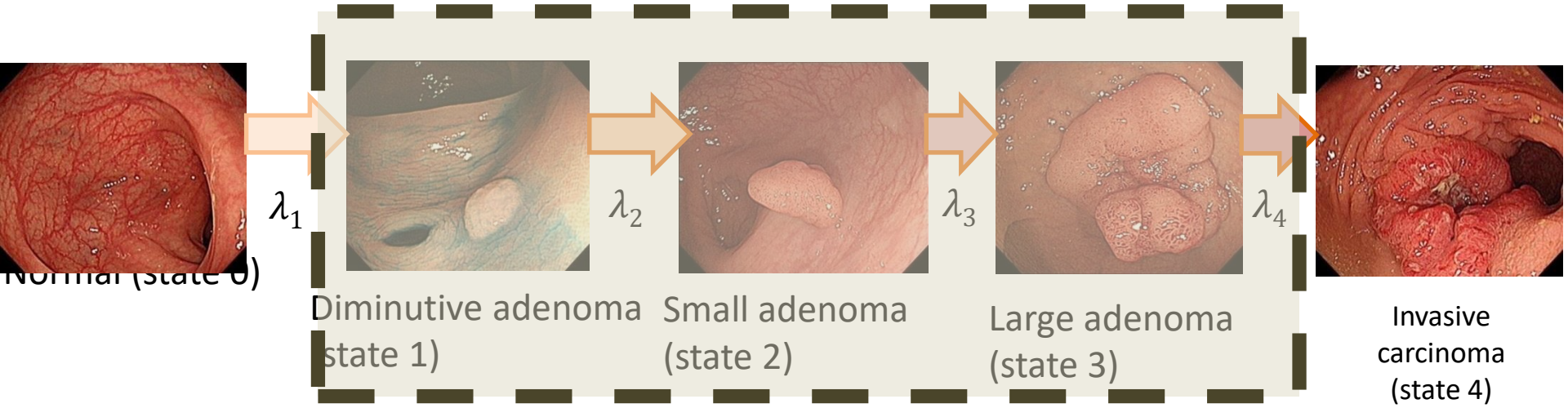
Stochastic model for non-standard case–cohort design

Tony Hsiu-Hsi Chen,*,† Ming-Fang Yen, Ming-Neng Shiu,
Tao-Hsin Tung and Hui-Min Wu

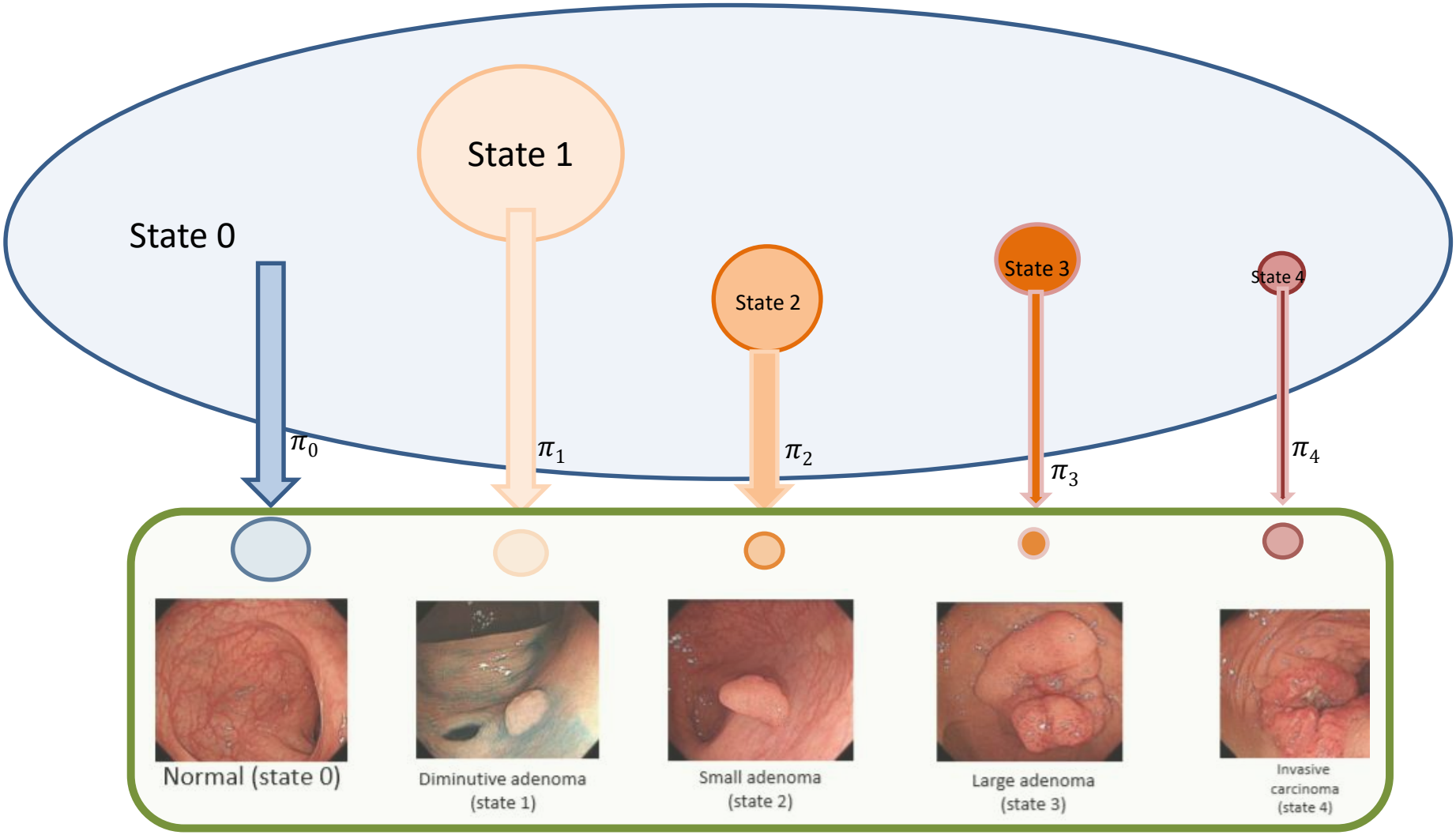
Institute of Preventive Medicine, College of Public Health, National Taiwan University,
Room 207, 19 Hsuchow Road, Taipei 100, Taiwan

1. Quantifying the rates of disease progression using **non-standard case-cohort design**

2. Assessing the **efficacy of polypectomy** according to disease characteristics



Statistics in Medicine *Statist. Med.* 2004; **23**:633–647 (DOI: 10.1002/sim.1610)



	Normal	(%)	Polyp	(%)	colorectal cancer	(%)
Cohort	10,496		2,652		760	
Sample	305	2.9	300	11.3	116	15.3

Disease progression and Sampling fraction

$$P(0 \to j, t_i \mid S=1)$$

$$= \frac{P(S=1 \mid 0 \to j, t_i) P(0 \to j, t_i)}{\sum_{j=1}^{n} P(S=1 \mid 0 \to j, t_i) P(0 \to j, t_i)} = \frac{\pi_j P(0 \to j, t_i)}{\sum_{j=1}^{n} \pi_j P(0 \to j, t_i)} = \frac{\pi_j P_{0j}(t_i)}{\sum_{j=1}^{n} \pi_j P_{0j}(t_i)}$$

Likelihood function

$$\prod_i \left(\frac{\pi_0 P_{00}(t_i)}{\sum_{j=0}^{4} \pi_j P_{0j}(t_i)} \right)^{n_{i0}} \left(\frac{\pi_1 P_{01}(t_i)}{\sum_{j=0}^{4} \pi_j P_{0j}(t_i)} \right)^{n_{i1}} \left(\frac{\pi_2 P_{02}(t_i)}{\sum_{j=0}^{4} \pi_j P_{0j}(t_i)} \right)^{n_{i2}} \left(\frac{\pi_3 P_{03}(t_i)}{\sum_{j=0}^{4} \pi_j P_{0j}(t_i)} \right)^{n_{i3}}$$
$$\times \left(\frac{\pi_4 P_{04}(t_i)}{\sum_{j=0}^{4} \pi_j P_{0j}(t_i)} \right)^{n_{i4}}$$

Efficacy=1-O/E

O: observed rate of disease progression

E: expected rate of disease progression

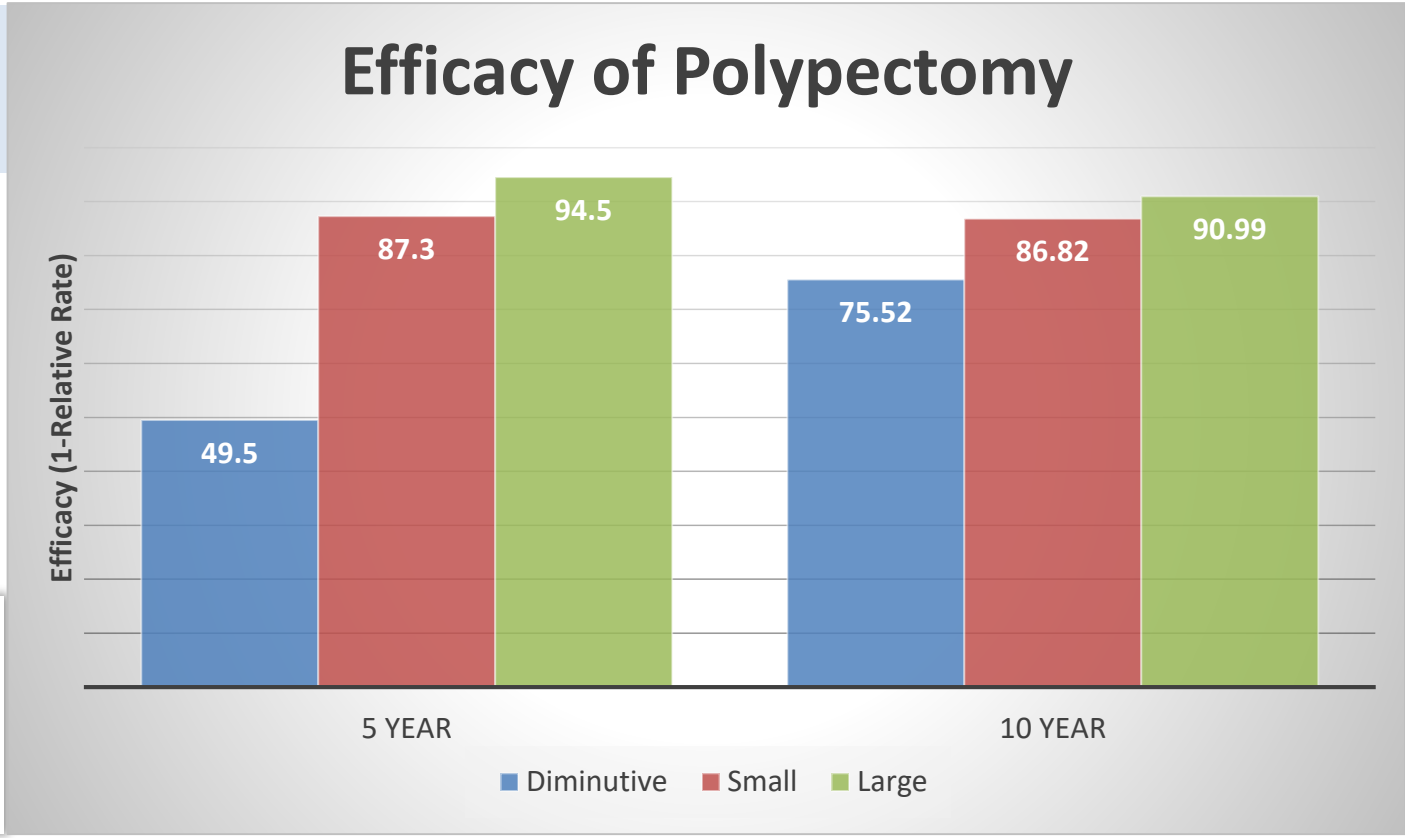
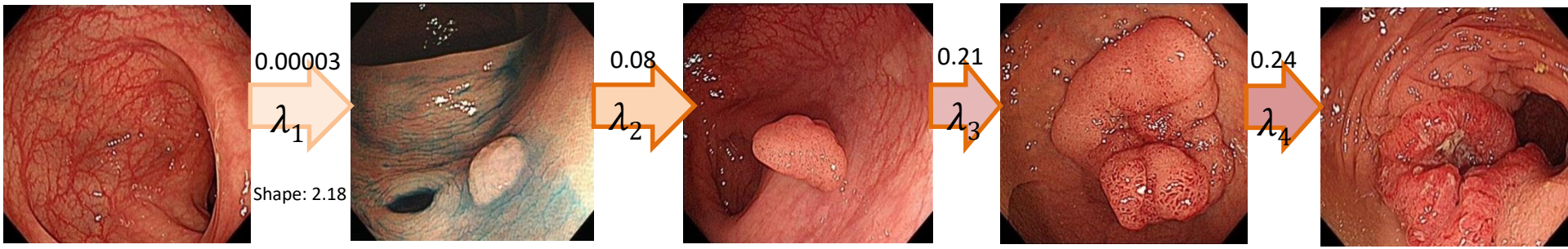


Table VI. Treatment efficacy of malignant transformation to colorectal cancer by adenoma size after 5-year and 10-year follow-up.

	Treatment efficacy	
Adenoma size	5 Year (per cent)	10 Year (per cent)
Diminutive	49.52	75.52
Small	87.31	86.82
Large	94.52	90.99



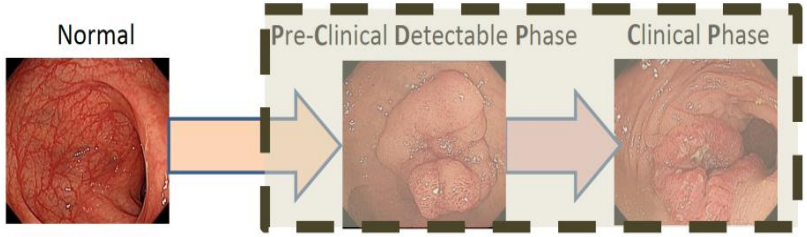
Multistate Markov Model with Two-stage Sampling Design

Aims

- Extent the **non-standard case-cohort design** from cross-sectional study to follow-up study
- Incorporation measurement error in the model
- Elucidate multistate disease progression
- Assess the state-specific effect of subject-specific characteristics with efficiency

Statistical Characteristics

- **Two-stage sampling design for multistate outcome**
  - Normal, **P**re-**C**linical **D**etectable **P**hase, **C**linical **P**hase



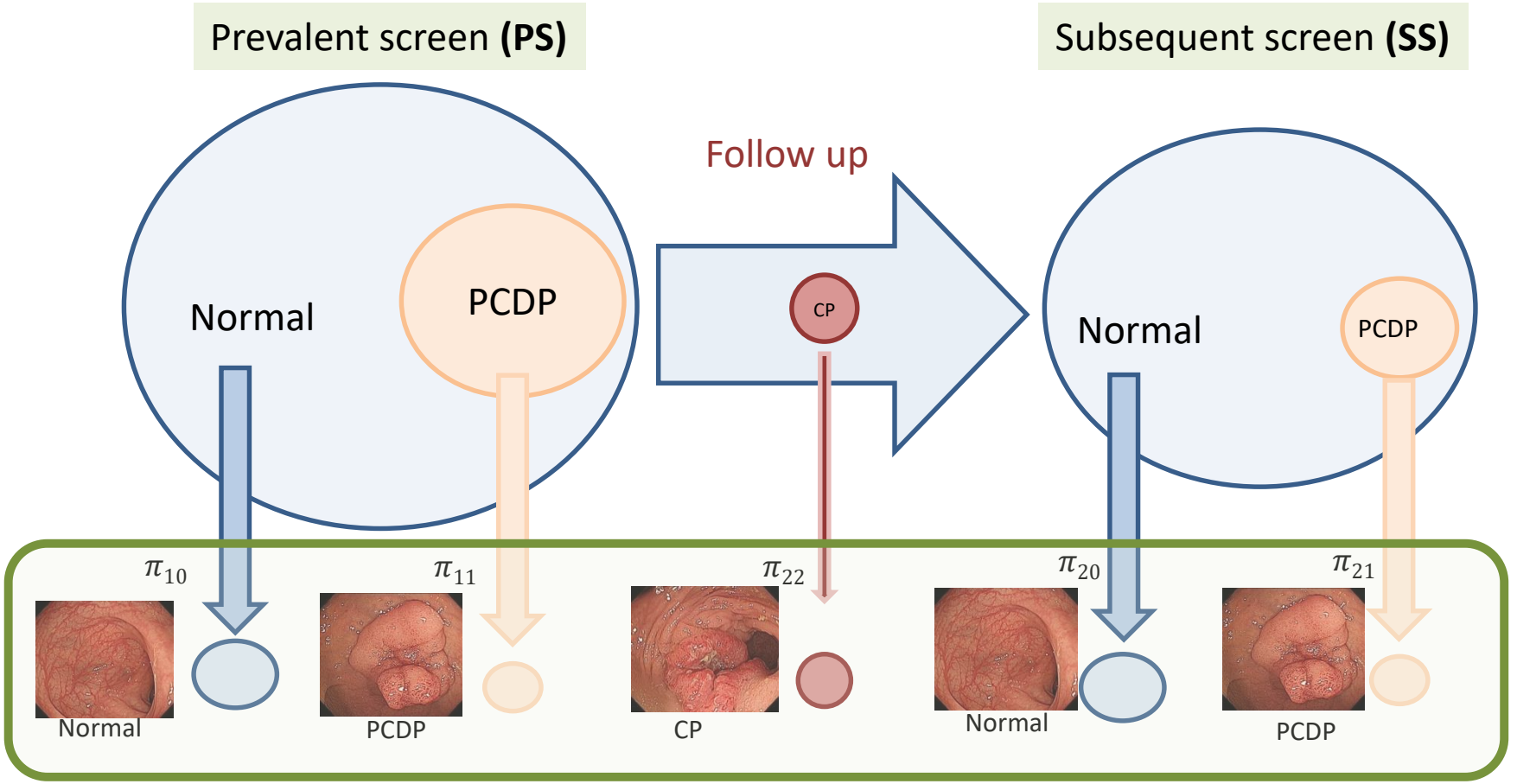
- **Incomplete information**
  - **Prevalent screening round**

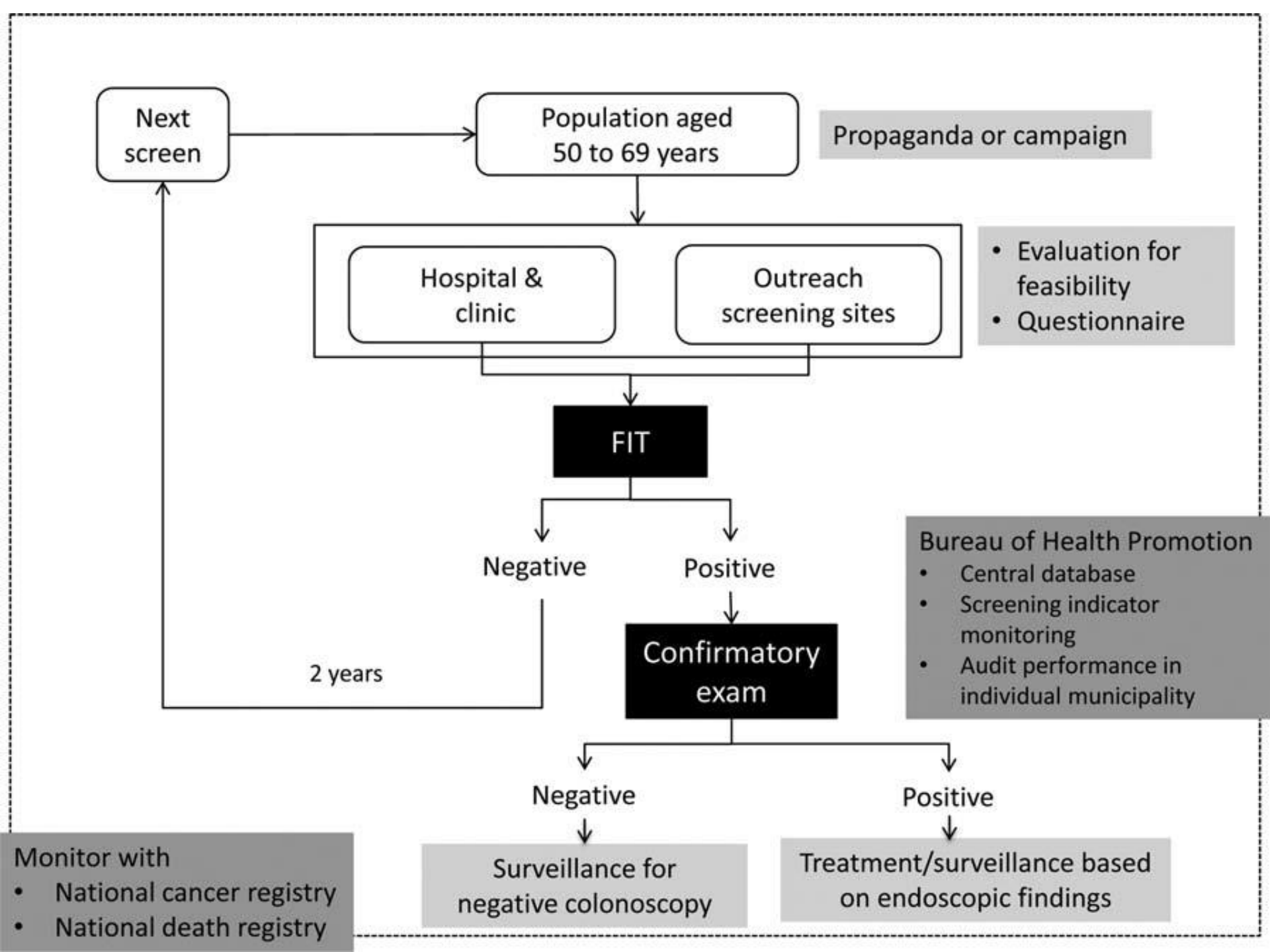
    Left Truncation
  - **Cases detected in subsequent screening round**

    Interval Censoring
  - **Surface to clinical phase between screen rounds (CP)**

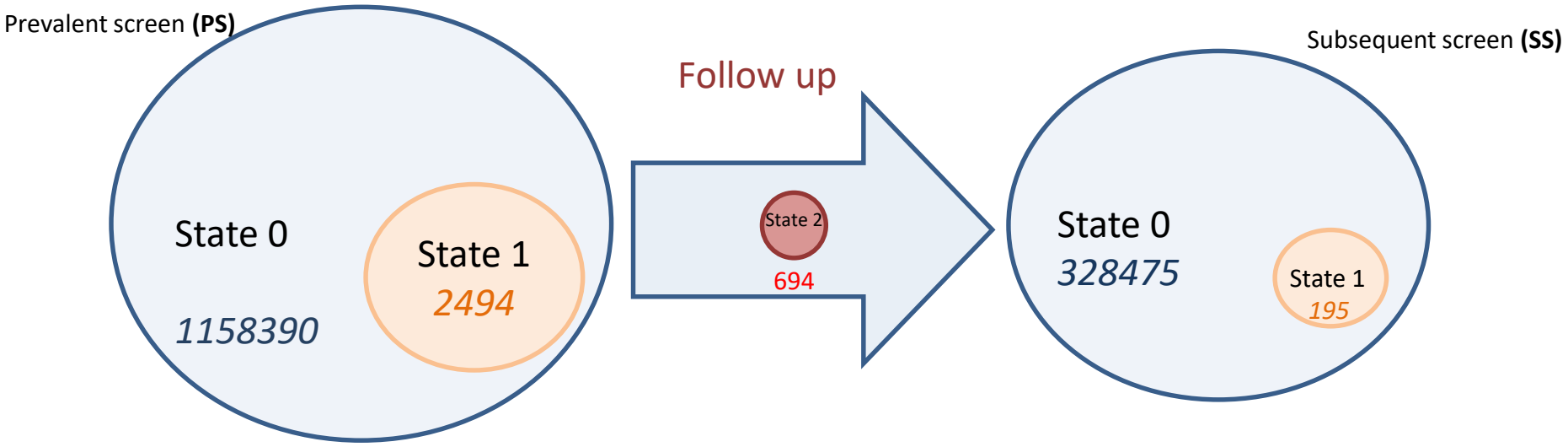
    Uncensored
- **Measurement error**

Two-stage sampling design for multistate screening data

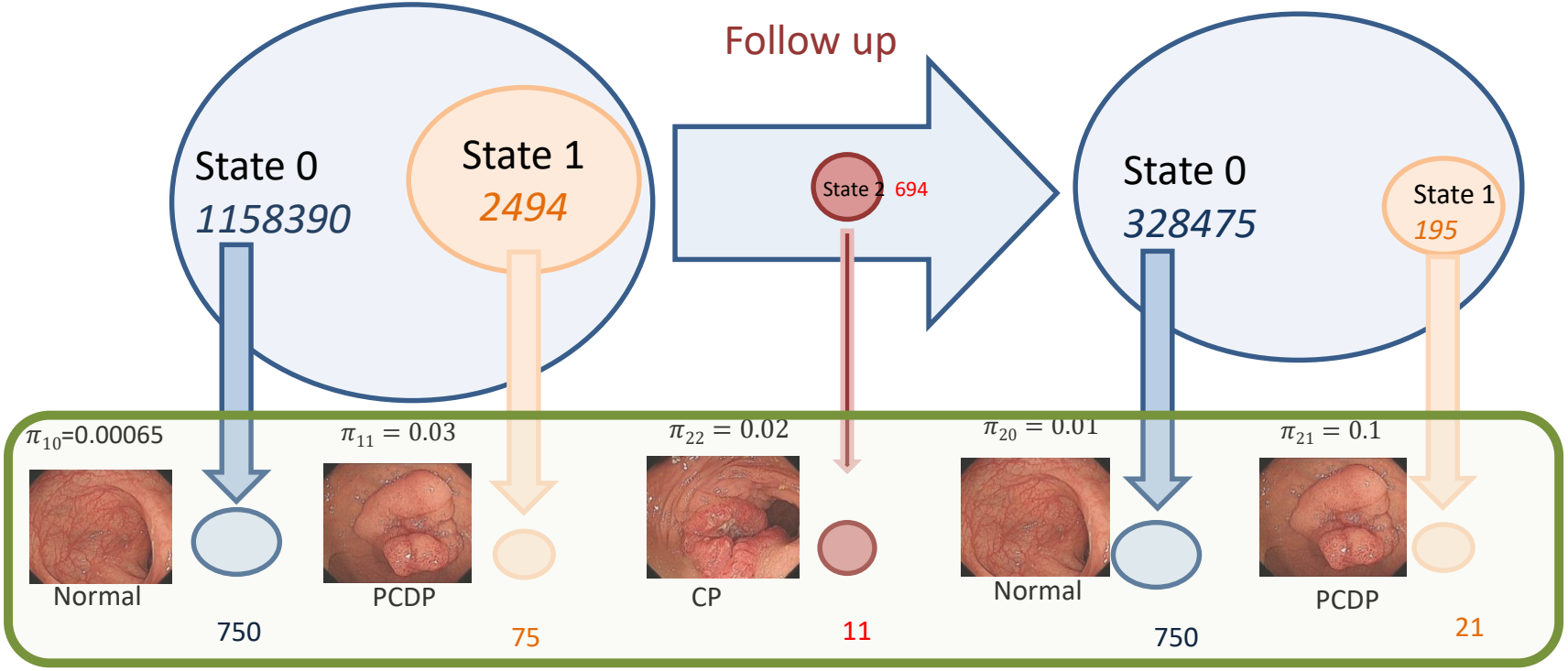



Next screen
Population aged 50 to 69 years
Propaganda or campaign
Hospital & clinic
Outreach screening sites
• Evaluation for feasibility
• Questionnaire
FIT
Negative
Positive
2 years
Bureau of Health Promotion
• Central database
• Screening indicator monitoring
• Audit performance in individual municipality
Confirmatory exam
Negative
Positive
Surveillance for negative colonoscopy
Treatment/surveillance based on endoscopic findings
Monitor with
• National cancer registry
• National death registry

Data on population-based screening for colorectal cancer in Taiwan



	Normal		PCDP		Interval cancer	
	Frequency	*(%)*	*Frequency*	*(%)*	*Frequency*	*(%)*
Prevalent screen						
Overall	1158390	(99.8)	2494	(0.2)		-
Male	444937	(38.4)	1347	(54.0)		-
Age ≧ 60	442793	(38.2)	1424	(57.1)		-
Subsequent screen						
Overall	328475	(99.7)	195	(0.06)	694	(0.21)
Male	110862	(33.8)	99	(50.8)	340	(49.0)
Age ≧ 60	168966	(51.4)	130	(66.7)	214	(30.8)



$$\lambda_1 = \lambda_{10} \exp\{\beta_{11} sex + \beta_{12}(age\ group)\},$$

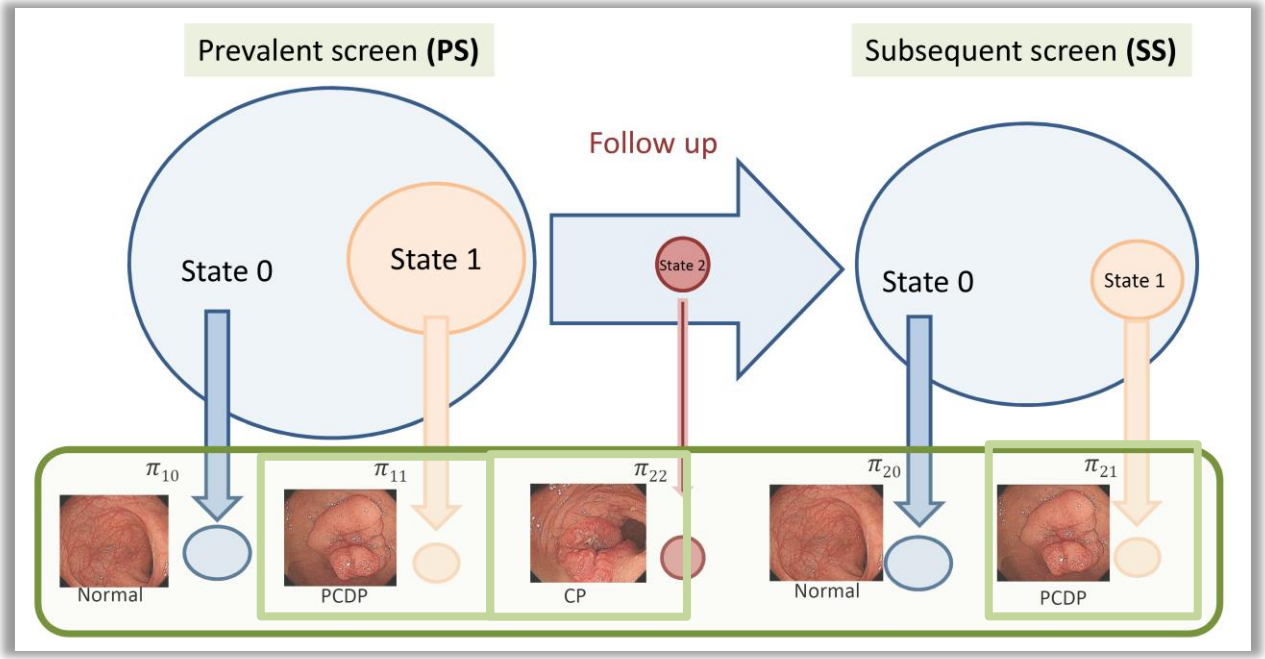
$$sex = \begin{cases} 1, \text{ if male} \\ 0, \text{ otherwise} \end{cases},$$

$$age\ group = \begin{cases} 1, \text{ if } \geq 60 \text{ years} \\ 0, \text{ otherwise} \end{cases}.$$

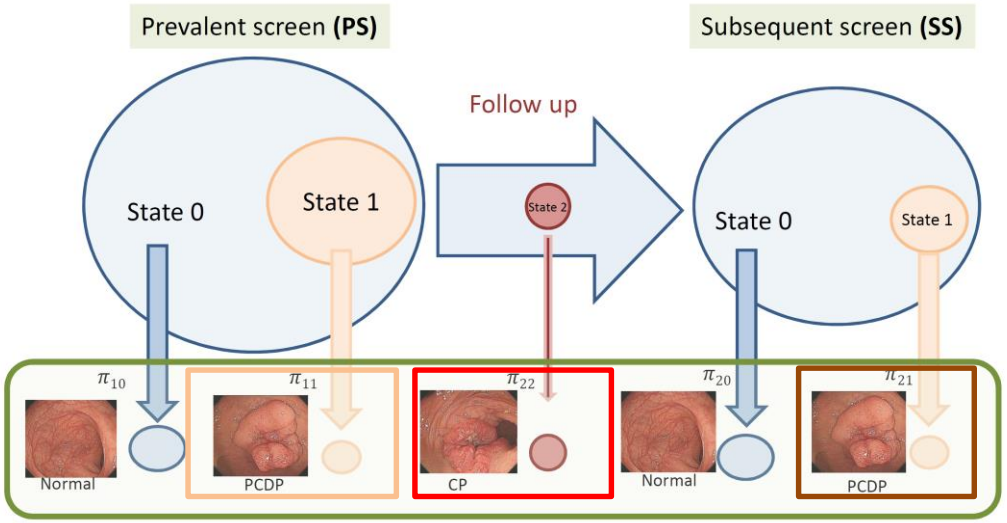
	Full date		Sampling Data set	
	Normal Subject in PS: 1158390		Normal subject in PS: 750 (π_{10}=0.00065)	
	Normal Subject in SS: 328475		Normal subject in SS: 750 (π_{20}=0.01)	
	PCDP case in PS: 2494		PCDP case in PS: 75 (π_{11}=0.03)	
	PCDP in SS: 195		PCDP case in SS: 21 (π_{21}=0.1)	
	Interval cancer cases: 694		Interval cancer cases: 11 (π_{22}=0.02)	
Parameter	**Estimate**	**(SD)**	**Estimate**	**(SD)**
λ_{10}	0.00043	(0.00004)	0.00057	(0.00013)
λ_{20}	0.32	(0.018)	0.33	(0.065)
β_1 (Male vs Female)	**0.51**	(0.076)	**0.54**	(0.200)
β_2 (Age >=60 vs <60)	**0.67**	(0.080)	**0.70**	(0.205)
Sensitivity	**0.81**	(0.026)	**0.82**	(0.081)

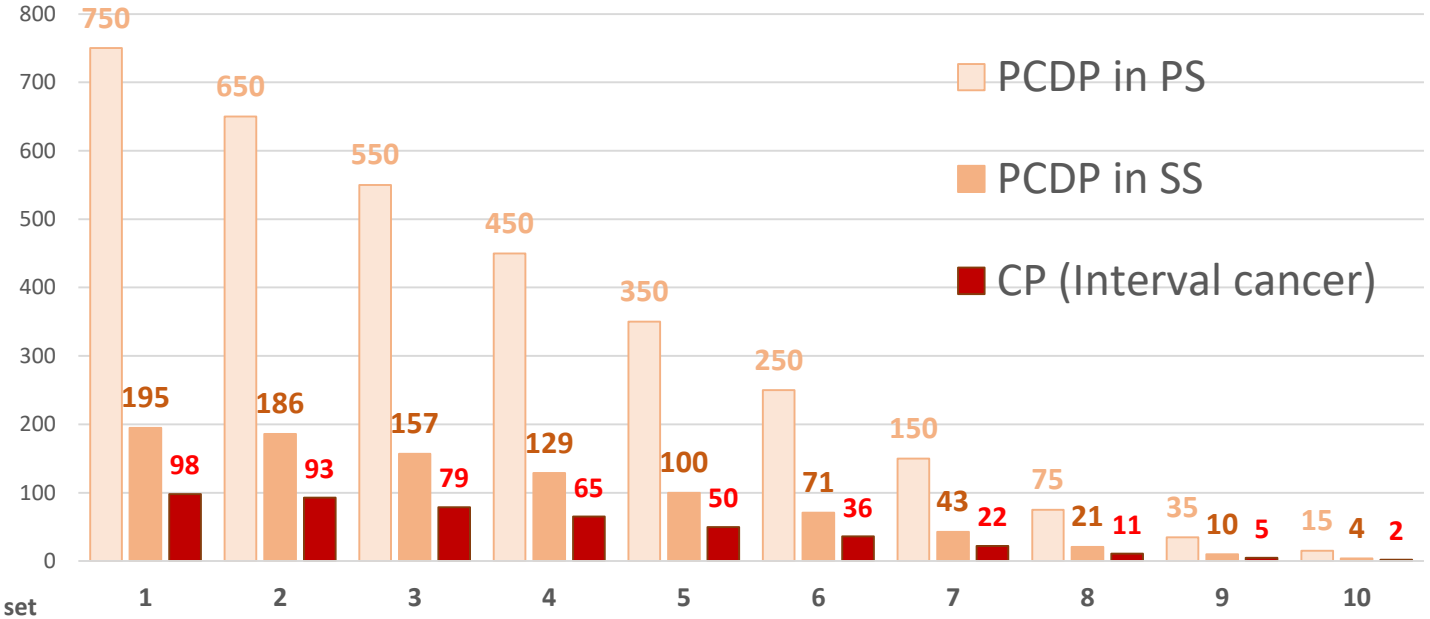
The influence of sample size on estimated effect of covariate

- Evaluating the effect of sample size on estimated results of covariate effect
  - Reduce sample size of screen detected cases (π_{11}, π_{21})
  - Reduce sample size of interval cancers (π_{22})



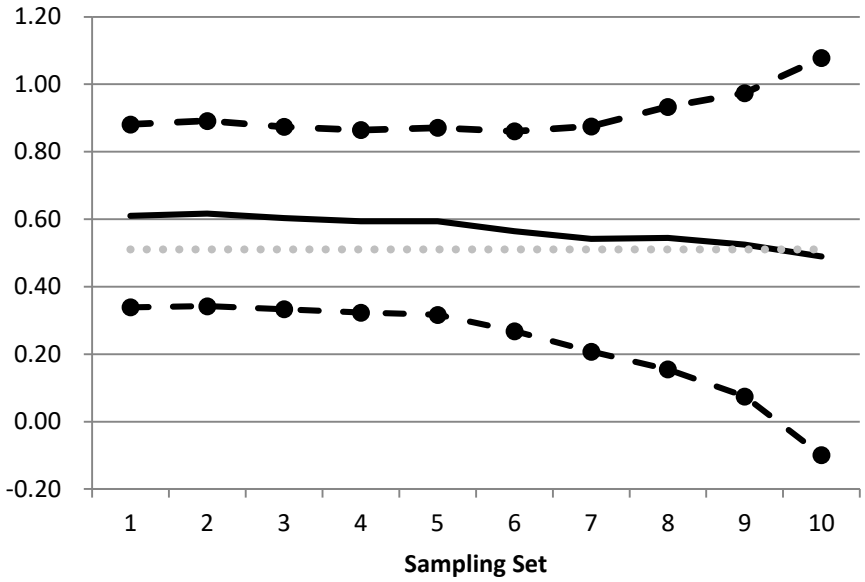
Normal subjects in PS: 750
(π_{10} =0.00065)
Normal subjects in SS: 1500
(π_{20} =0.0005)

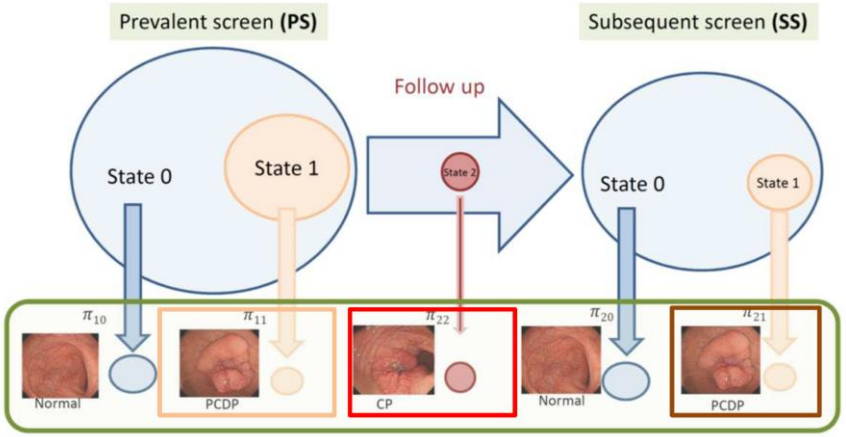




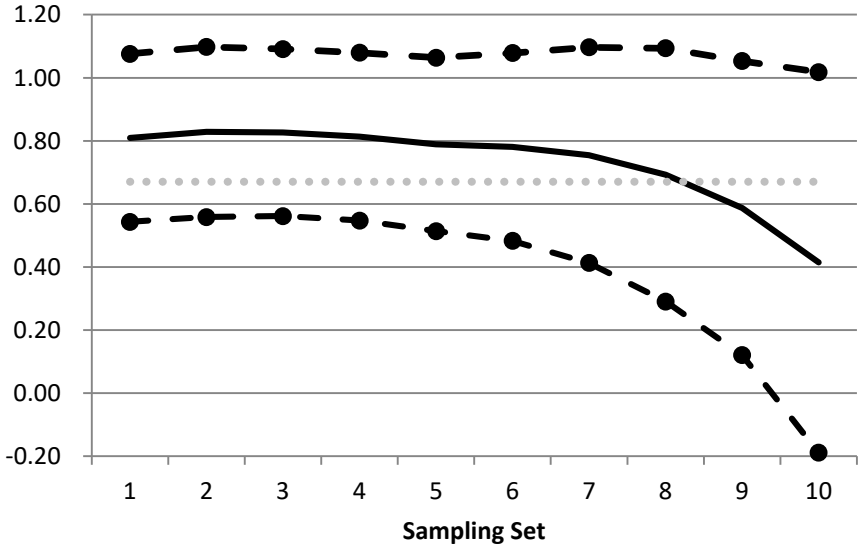
Sampling fraction										
	PCDP, PS (π_{11})	0.30	0.26	0.22	0.18	0.14	0.1	0.06	0.03	0.014	0.006
	PCDP, SS (π_{21})	1	0.95	0.81	0.66	0.51	0.36	0.22	0.11	0.05	0.02
	CP (π_{22})	0.14	0.13	0.11	0.09	0.07	0.05	0.03	0.016	0.007	0.003

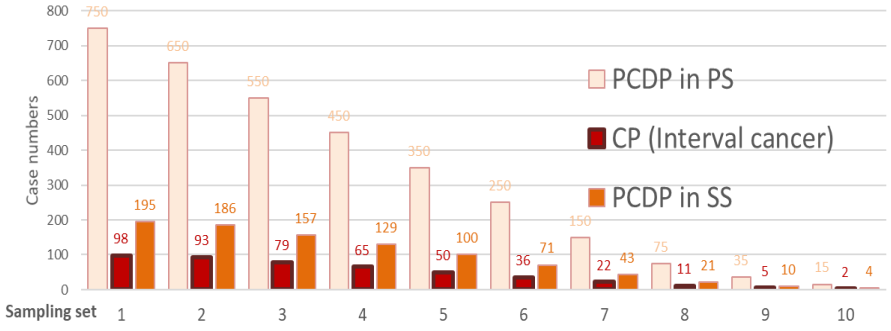
Sex


1.20
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
1 2 3 4 5 6 7 8 9 10
Sampling Set

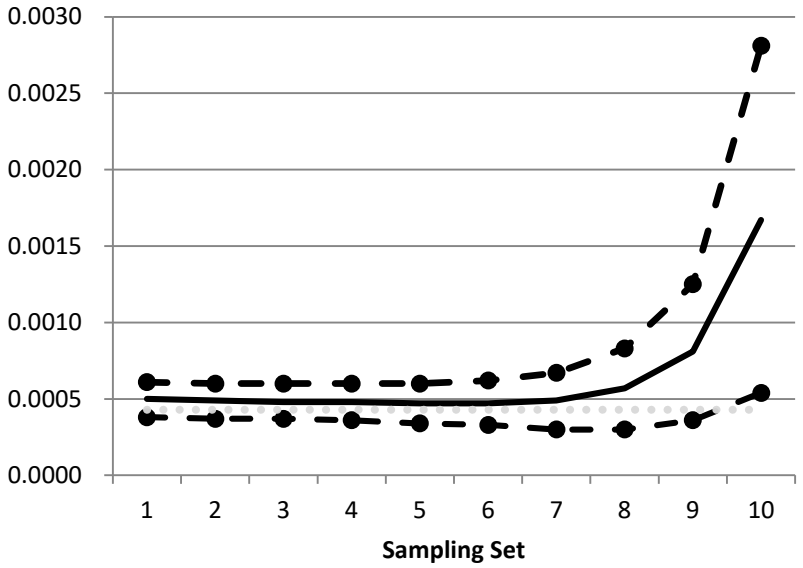

Prevalent screen (PS)
Subsequent screen (SS)
Follow up
State 0
State 1
State 2
State 0
State 1
π_{10}
π_{11}
π_{22}
π_{20}
π_{21}
Normal
PCDP
CP
Normal
PCDP

Age

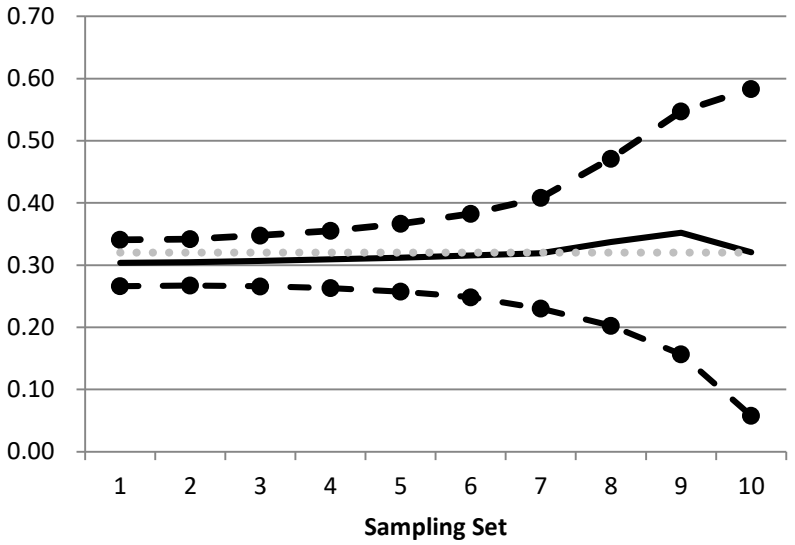

1.20
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
1 2 3 4 5 6 7 8 9 10
Sampling Set


Case numbers
800
700
600
500
400
300
200
100
0
Sampling set
1 2 3 4 5 6 7 8 9 10
750 650 550 450 350 250 150 75 35 15
98 195 93 186 79 157 65 129 50 100 36 71 22 43 11 21 5 10 2 4
PCDP in PS
CP (Interval cancer)
PCDP in SS

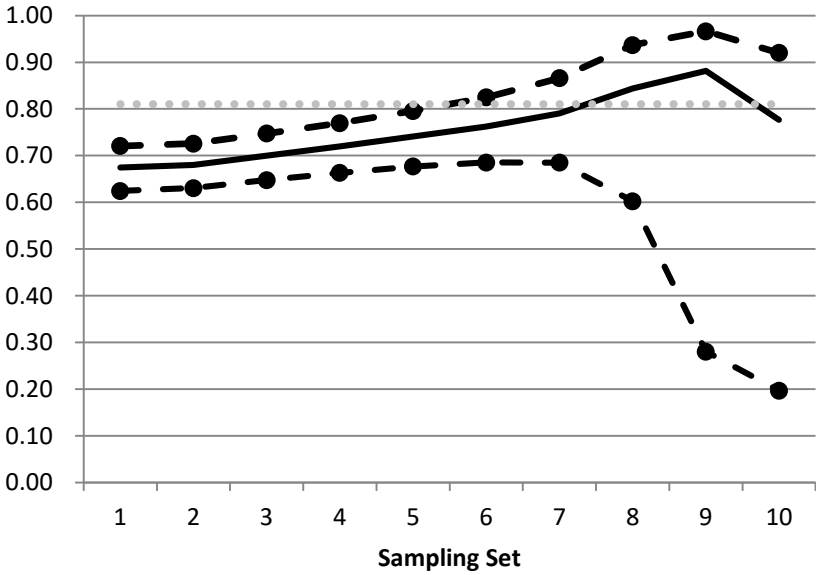
Incidence rate

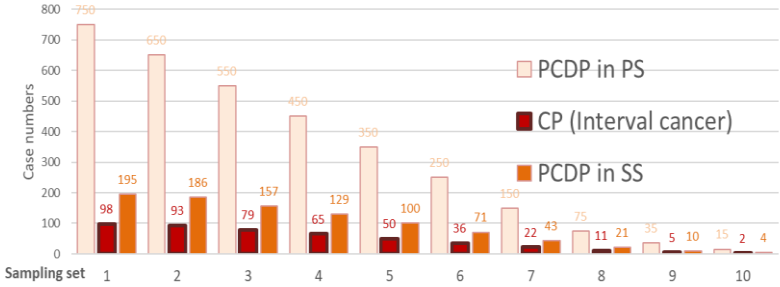


Progress rate

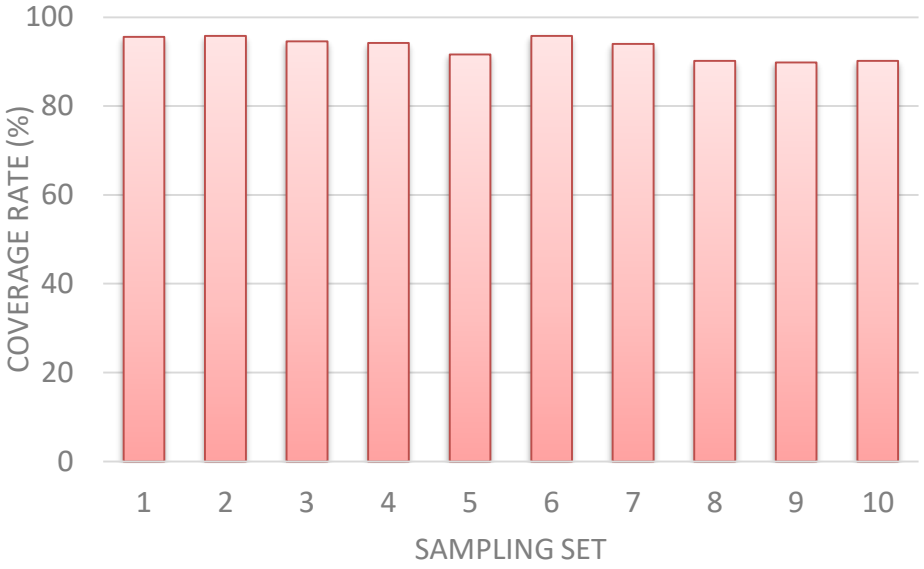


Sensitivity

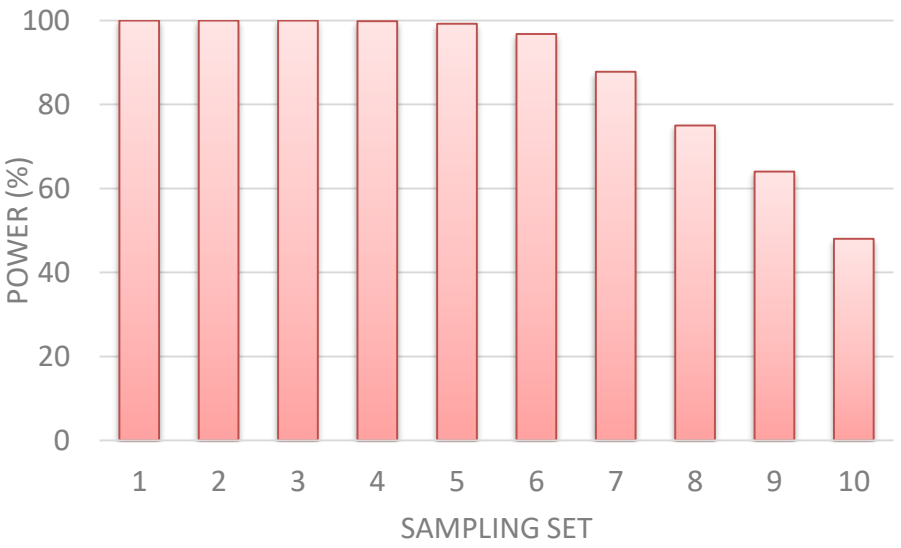




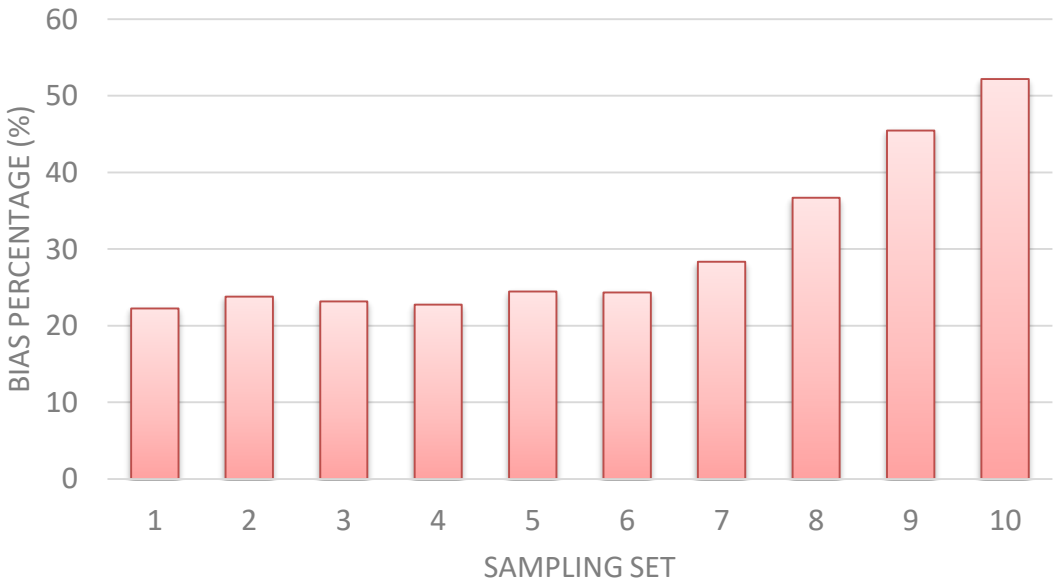
Coverage rate, Sex

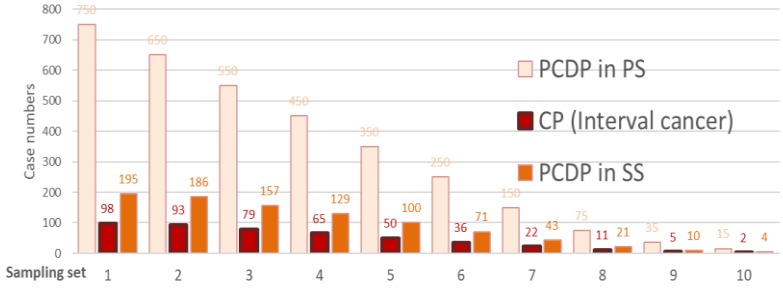


Statistical power, Sex

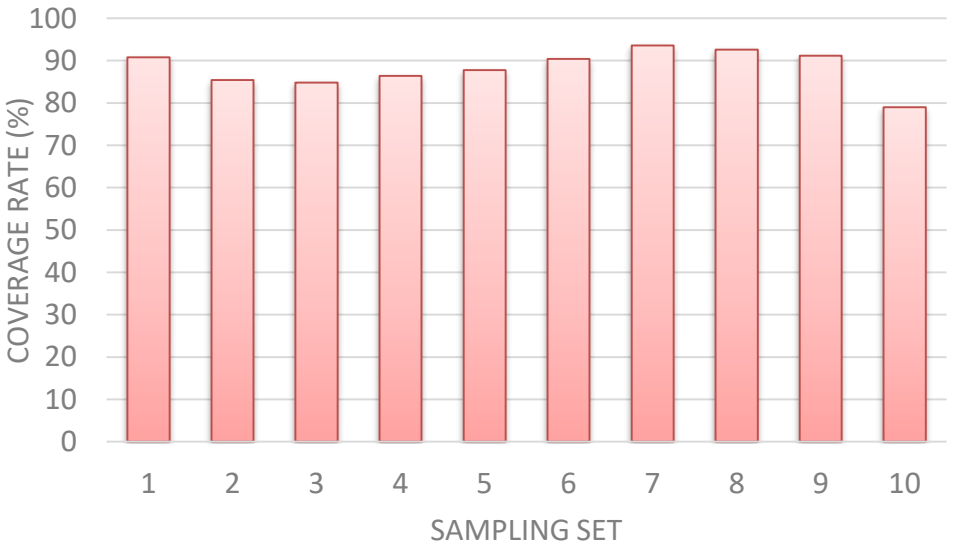


Bias percentage, Sex

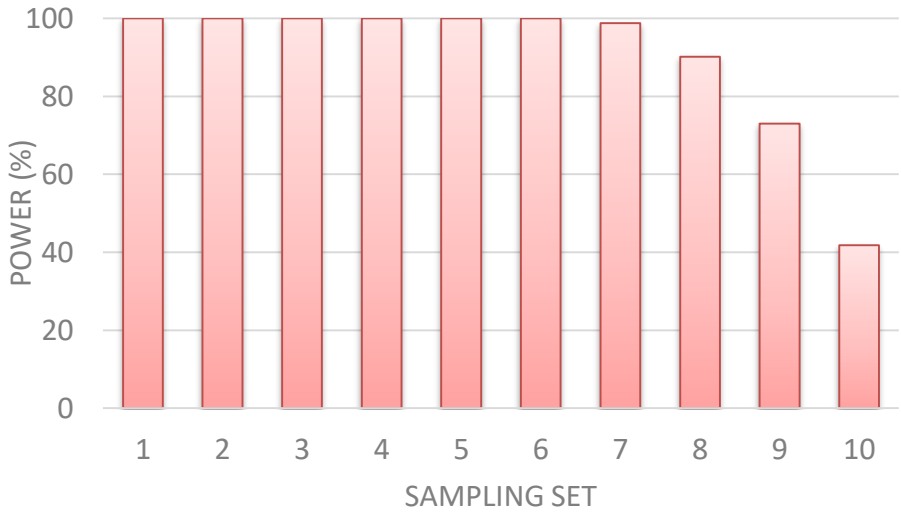




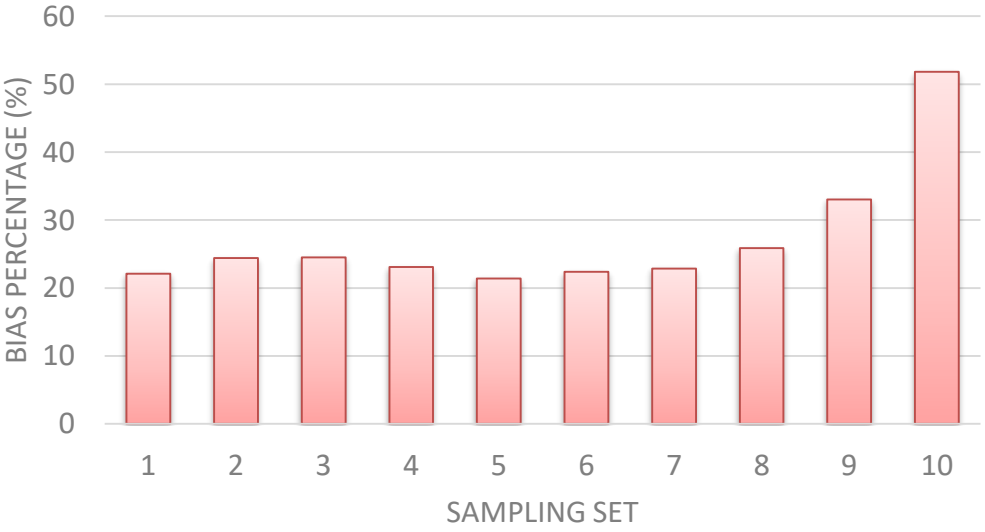
Coverage rate, Age

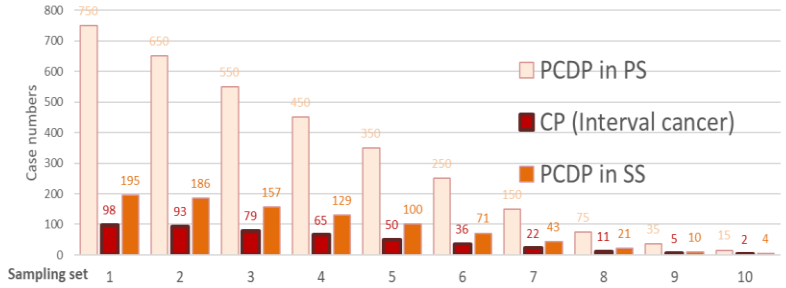


Statistical power, Age



Bias percentage, Age





Summary

- Two-stage sampling design with multistate outcome
  - Extend the cross-sectional non-standard case-cohort design to longitudinal follow-up data
  - Incorporating measurement error
  - Accommodate the incomplete information (truncation and censoring) characteristics of data

- Assess the state-specific effect of covariates based on sampled data