Multi-state Model for the Estimation of Overdiagnosis in the Breast Cancer Screening Program

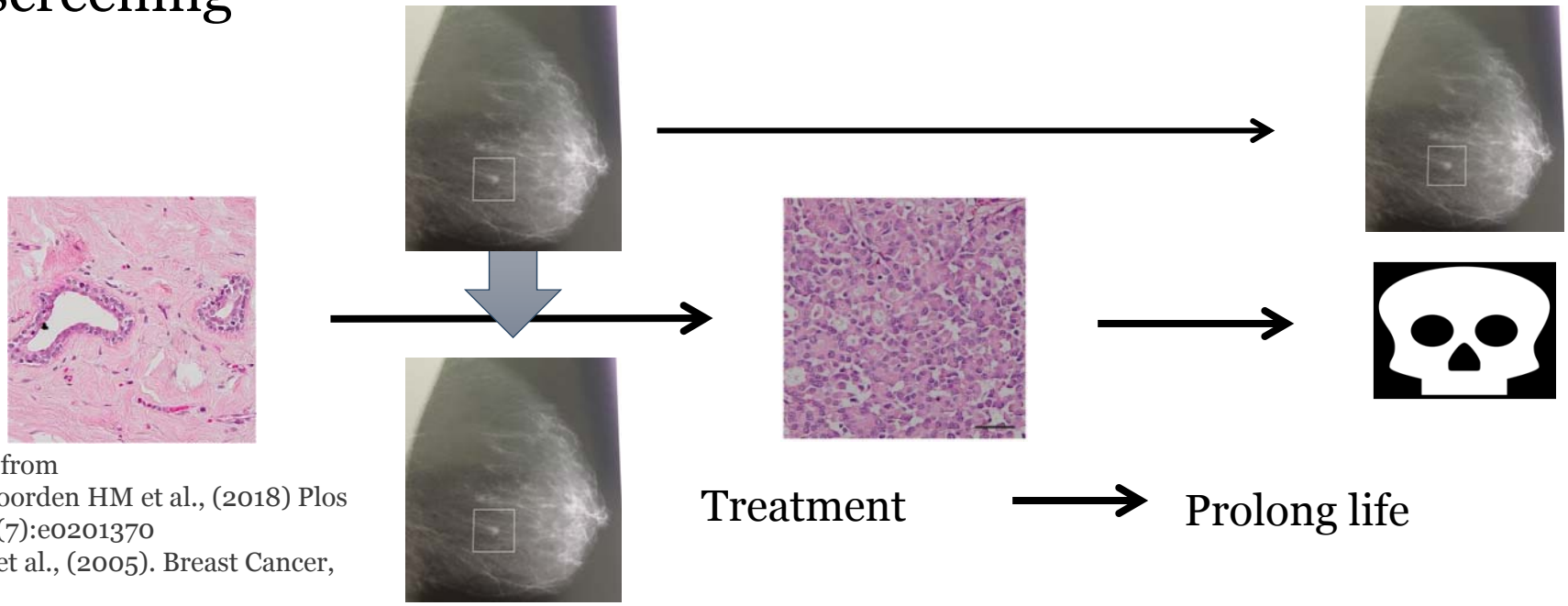
Wendy Yi-Ying Wu

Department of Radiation Sciences, Oncology

UMEÅ UNIVERSITY

Overdiagnosis

Screening detects a breast cancer that would not have presented clinically in the woman's lifetime in the absence of screening



Pictures from
Duivenvoorden HM et al., (2018) Plos ONE, 13(7):e0201370
Tabar L et al., (2005). Breast Cancer, Thieme

Qantification is difficult!

Estimates varies from 0-56% (EUROSCREEN)

Control group?

Evaluation method?

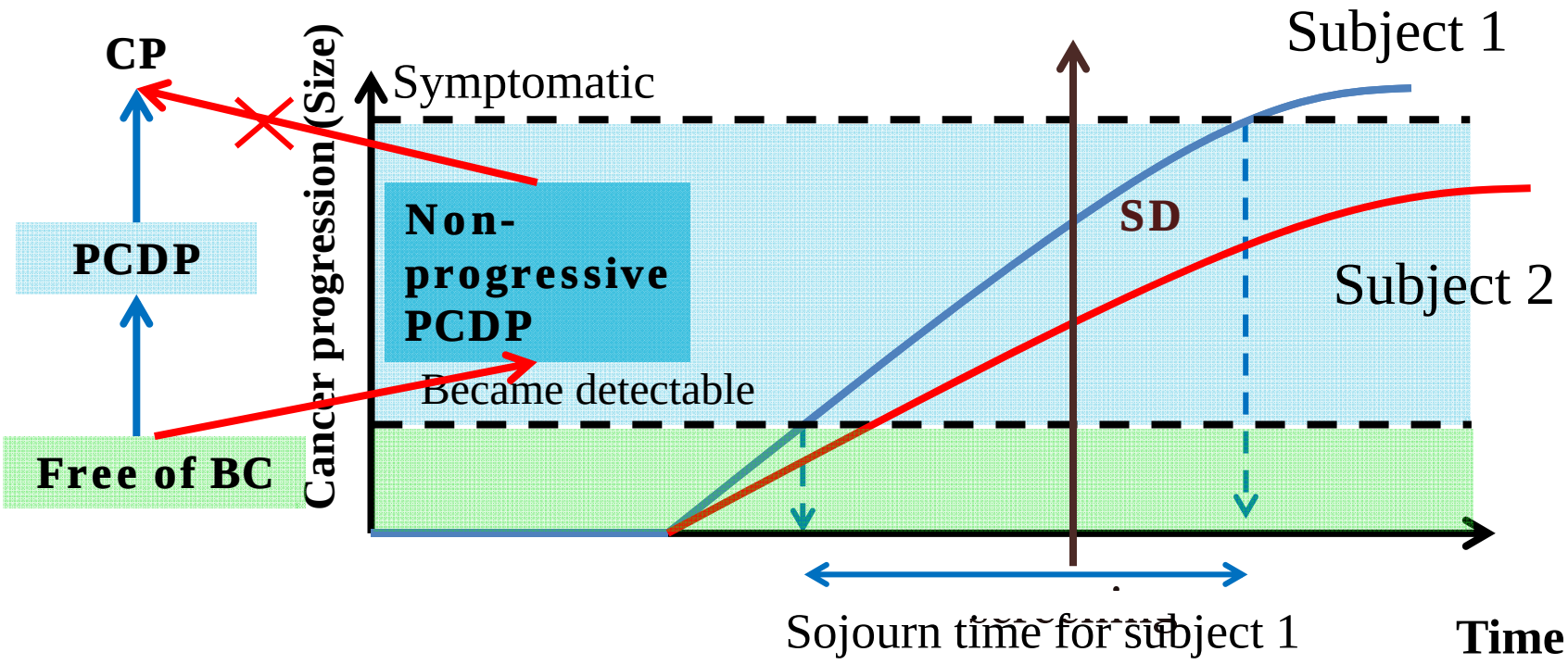
Useful method for service-screening?

Definition?

Lead time adjustment?

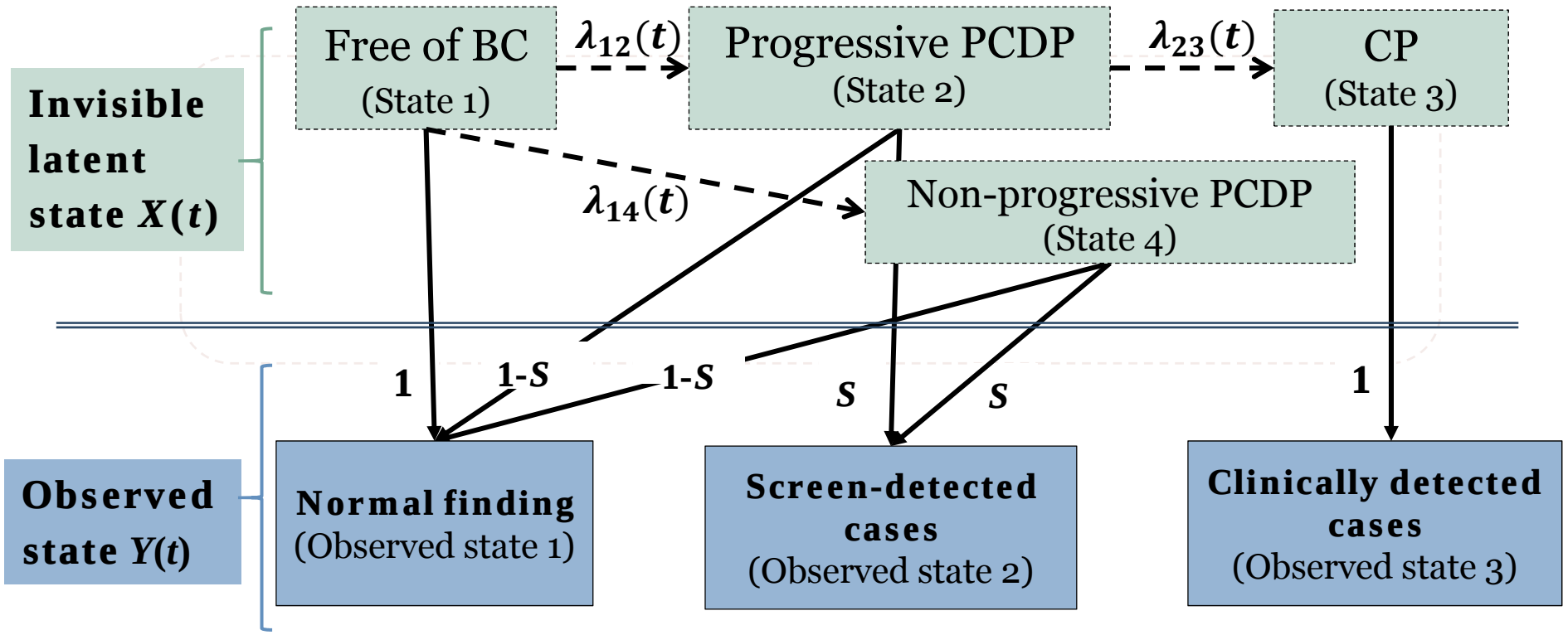
Denominators?

Illustration



BC: Breast cancer, PCDP: Preclinical screen-detected phase, CP: Clinical phase,
SD: Screen-detected case

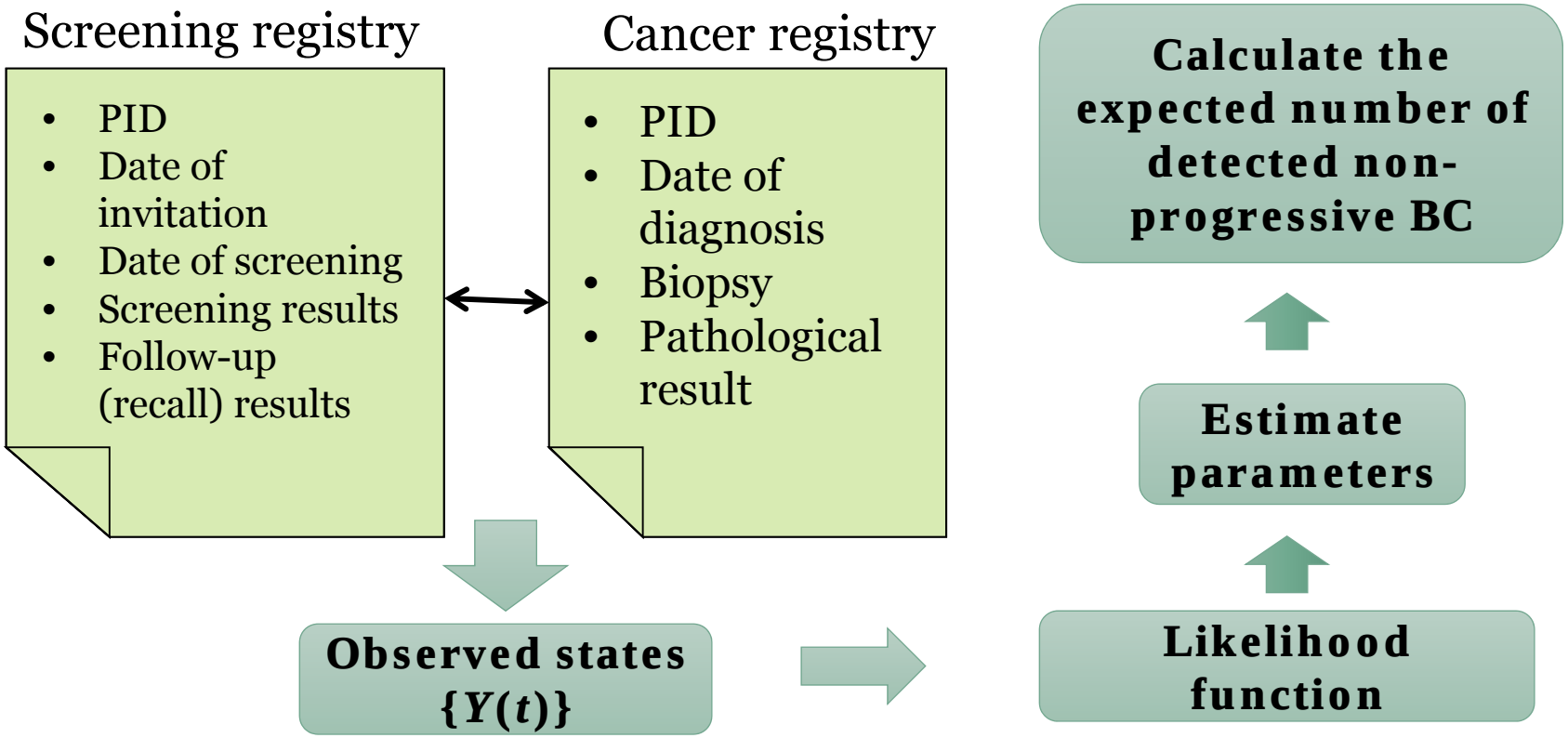
Multi-state model



$\lambda_{ij}(t)$: transition rate from state i to state j at time t, S: test sensitivity

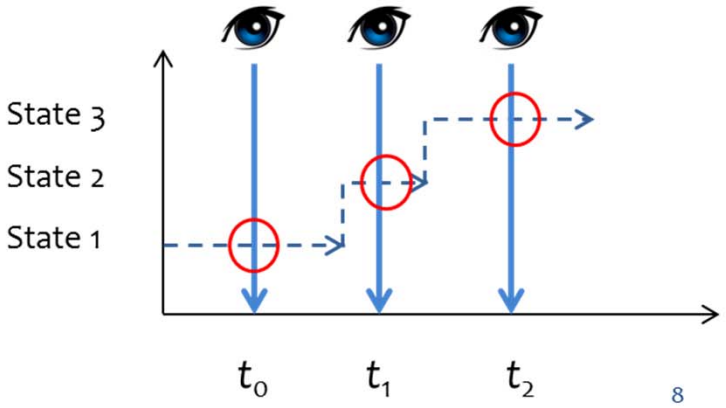
BC: breast cancer, PCDP: preclinical screen-detectable phase, CP: clinical phase

Data sources and work process



Screening Data

- Panel data (longitudinal follow-up data)
  - State only observed at a finite number of times (interval censoring)
  - Don't know the state between times



Let $Y(t)$ denote the state occupied at time t by a random chosen individual

$$\{Y(t_0) = 1, Y(t_1) = 2, Y(t_2) = 3\}$$

Screening Data

- Left truncation
  - Only the asymptomatic subjects will be invited
- Time-inhomogeneous model
  - The incidence rate depends on the age
- Measurement error
  - Sensitivity ≠ 100%
- Informative censoring ??
  - The subjects with shorter time to clinical phase tend to be clinically detected if the participation rate is low

Conditional probability

Piece-wise rate/ parametric

Hidden Markov model

EM algorithm??

Supplementary Table 2. The Observed Data and Corresponding Likelihood in The Screened and Control Groups by Modes of Detection

Type	Observed states	Likelihood[a]
Screened group		
SD cases at k^{th} round	$\{Y(t_1)=2 \mid X(t_1)\neq 3\}$ for k=1, and $\{Y(t_k)=2, Y(t_{k-1})=1,\ldots,Y(t_1)=1 \mid X(t_1)\neq 3\}$ for $1<k\leq m$	$\frac{\left(\prod_{\ell=0}^{k-2} P_{11}(t_{\ell-1},t_\ell)\right)}{1-P_{13}(t_0,t_1)} \times \{P_{11}(t_{k-2},t_{k-1}) \times (P_{12}(t_{k-1},t_k)+P_{14}(t_{k-1},t_k)) \times S + (P_{12}(t_{k-2},t_{k-1}) \times P_{22}(t_{k-1},t_k)+P_{14}(t_{k-2},t_{k-1})) \times (1-S)\}$
CD cases between k^{th} and $k+1^{th}$ round	$\{Y(t)=3, Y(t_k)=1,\ldots,Y(t_1)=1 \mid X(t_1)\neq 3\}$ for $1\leq k<m$, and $t_k<t<t_{k+1}$.	$\frac{\left(\prod_{\ell=0}^{k-1} P_{11}(t_{\ell-1},t_\ell)\right)}{1-P_{13}(t_0,t_1)} \times \{P_{11}(t_{k-1},t_k) \times P_{12}(t_k,t) \times \lambda_{23}(t) + P_{12}(t_{k-1},t_k) \times (1-S) \times P_{22}(t_k,t) \times \lambda_{23}(t)\}$
NF women in the last round of screening	$\{Y(t_m)=1, Y(t_{m-1})=1,\ldots,Y(t_1)=1 \mid X(t_1)\neq 3\}$	$\frac{\left(\prod_{\ell=0}^{m-1} P_{11}(t_{\ell-1},t_\ell)\right)}{1-P_{13}(t_0,t_1)} \times \{P_{12}(t_{m-1},t_m) \times (1-S)+P_{11}(t_{m-1},t_m)\}$
Control group		
CD	$\{Y(t)=3 \mid X(t_1)\neq 3\}$ for $t_1<t\leq t_m$	$\frac{1}{1-P_{13}(t_0,t_1)} \times \{P_{11}(t_0,t_1) \times P_{12}(t_1,t) \times \lambda_{23}(t)+P_{12}(t_0,t_1) \times P_{22}(t_1,t) \times \lambda_{23}(t)\}$
Non-BC cases	$\{Y(t_m)\neq 3 \mid X(t_1)\neq 3\}$	$\frac{1}{1-P_{13}(t_0,t_1)} \times \{P_{11}(t_0,t_1) \times (1-P_{13}(t_1,t_m))+P_{12}(t_0,t_1) \times P_{22}(t_1,t_m)+P_{14}(t_0,t_1)\}$

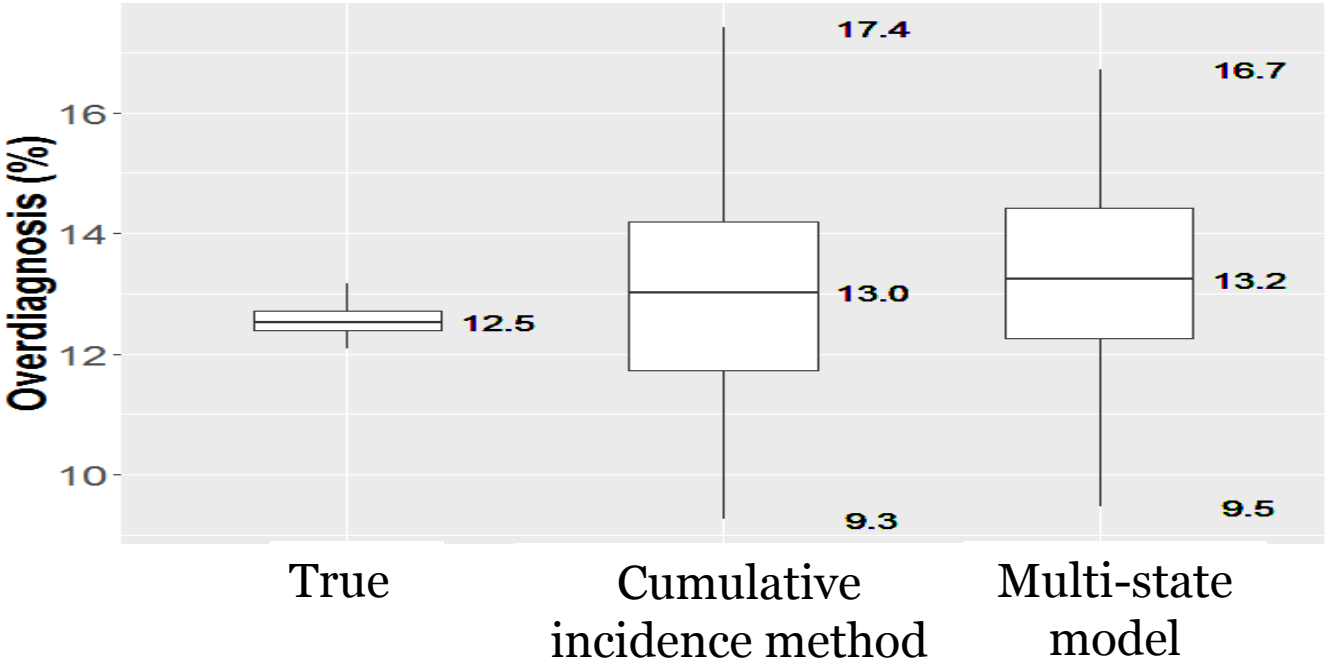
[a] for $\ell\leq 0$, $P_{1j}(t_{\ell-1},t_\ell)=1$ if $j=1$, otherwise $P_{1j}(t_{\ell-1},t_\ell)=0$

BC: breast cancer; CD: clinically detected; m; the total number of screening; NF: negative finding; SD: screen-detected cases; t_k : the time at k^{th} round of screening; $X(t)$: invisible latent state; $Y(t)$: observed state

Estimation of overdiagnosis in breast cancer screening using a non-homogeneous multi-state model: A simulation study

J Med Screen
0(0) 1–8

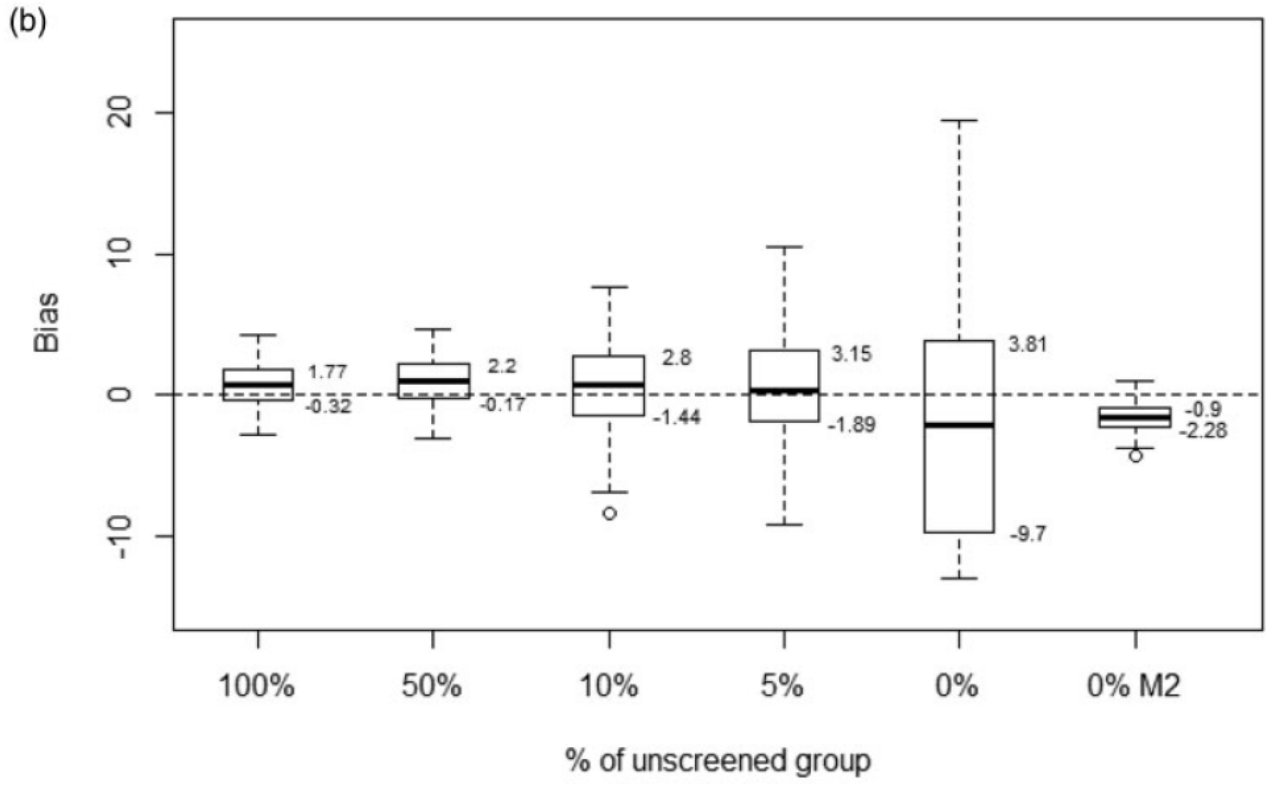
DOI: 10.1177/0969141317733294
journals.sagepub.com/home/msc
SAGE



Randomized controlled trial

1: 1 ratio screened and control, 100% participation rate

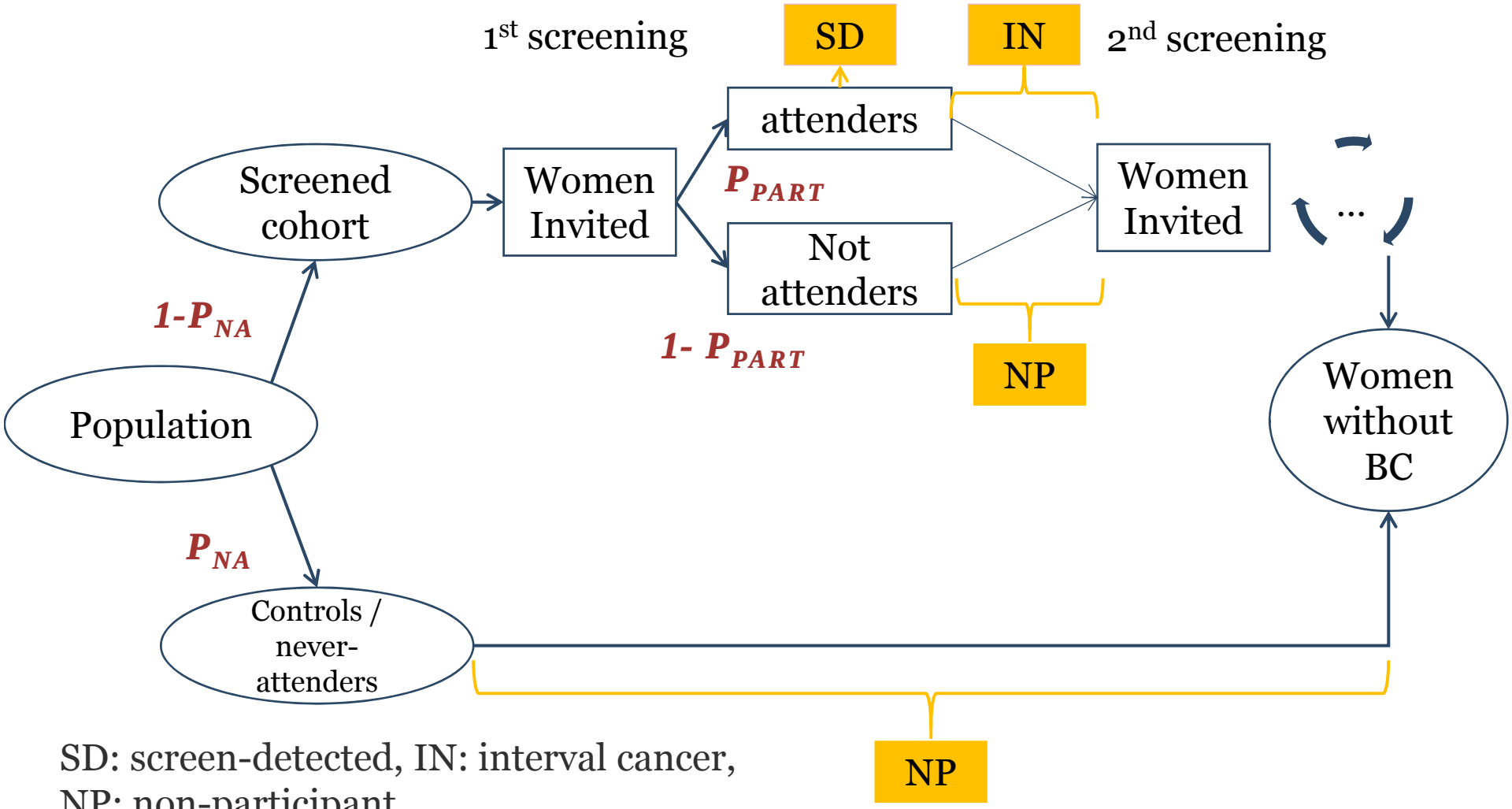
Conclusion: Both methods can provide a proper estimate



Control group provided information to stabilize the model.
(help to deal with the identifiability problem)

Control group in the service screening program?

- **Never-attenders** (who were invited but never participated in the screening program) can provide the similar information under certain conditions, i.e. similar incidence rate
  - Group A: those who will never attend any round of screening
  - Group B: those who might attend the screening but does not participate in the current/previous rounds
    - Shorter time to CP → few rounds of invitation→ more clinical cases (**observation process and disease process are not independent= informative censoring??)**

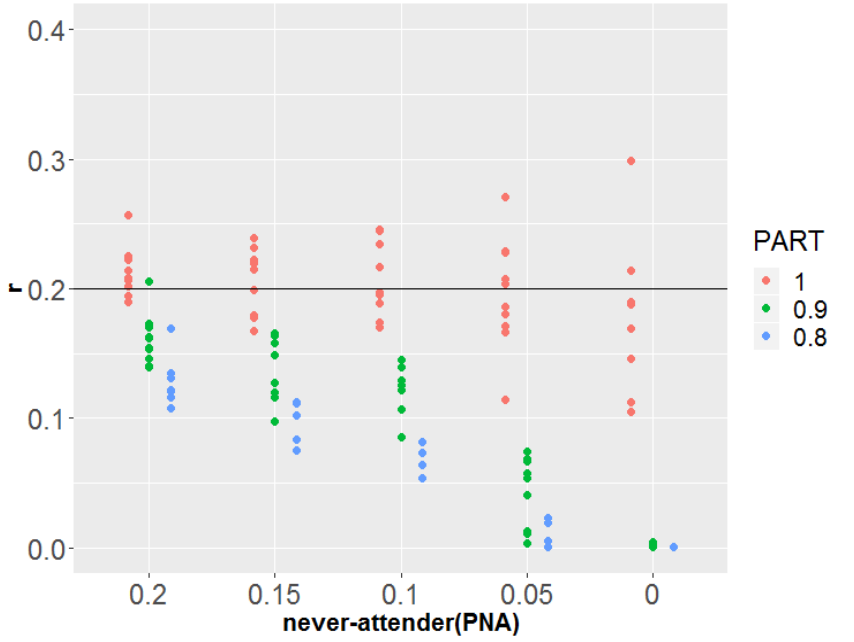


SD: screen-detected, IN: interval cancer,
NP: non-participant

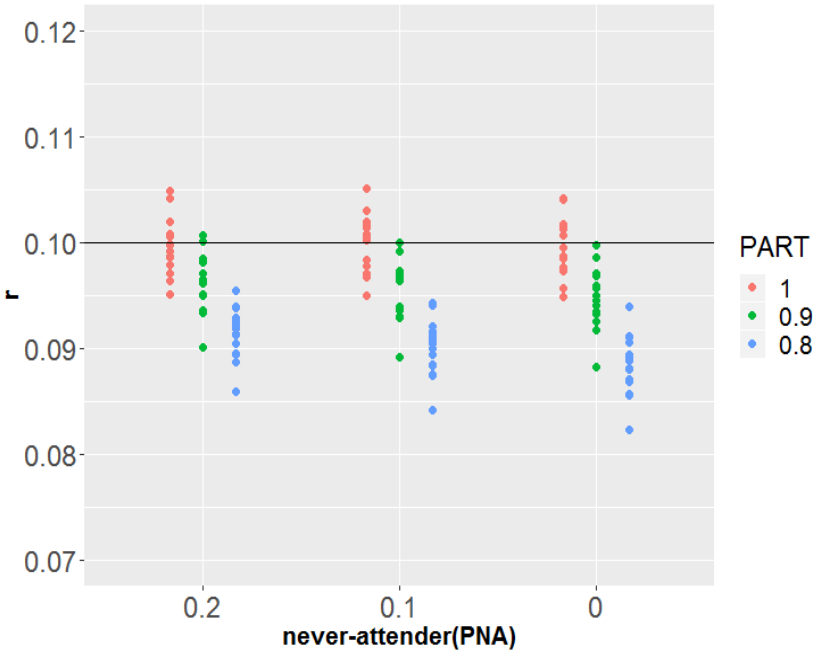
$$r = \frac{\lambda_{14}(t)}{\lambda_{12}(t)}$$

Simulation results

Non-homogeneous model

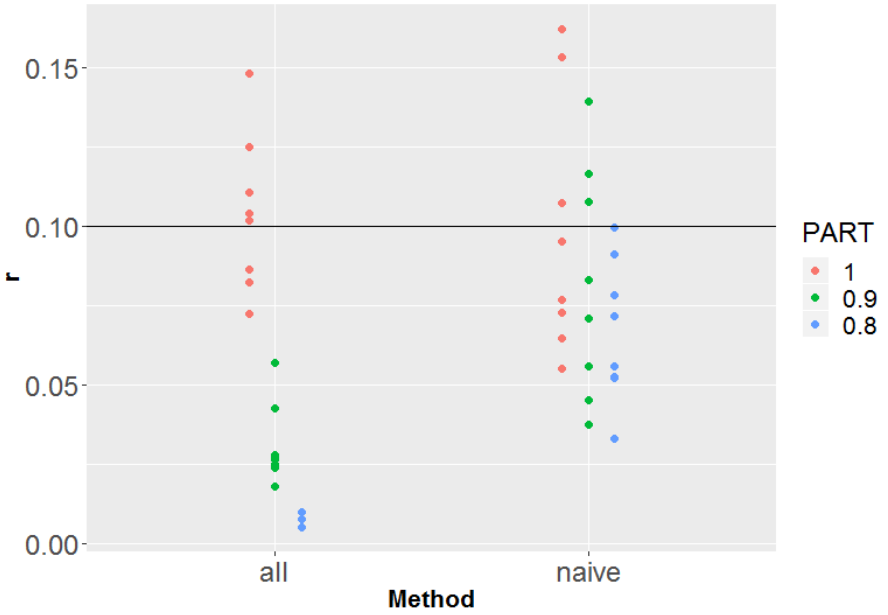


Homogeneous model

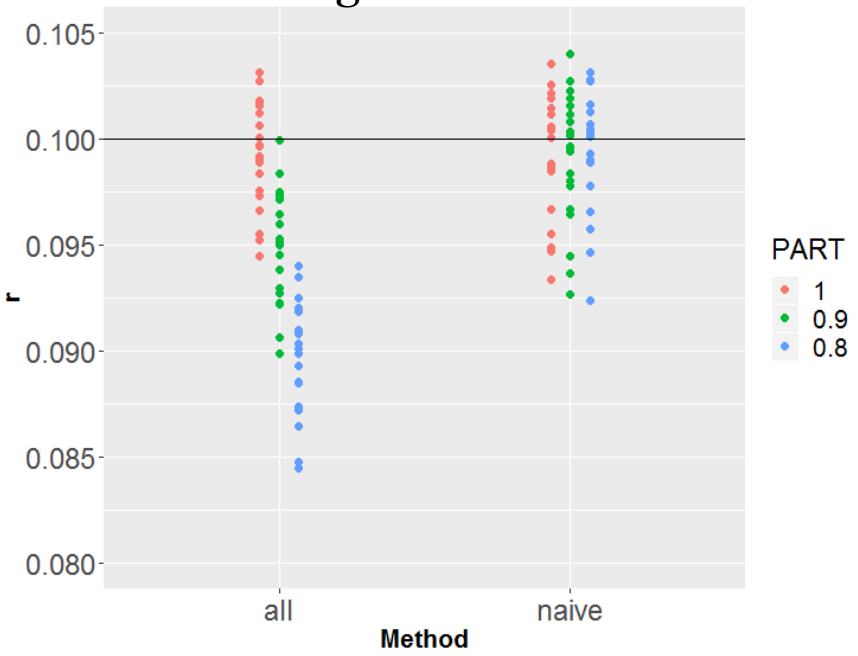


A naive method (removing all the never-attenders)

Non-homogeneous model



Homogeneous model



Given on P_{NA}=0.1

Future studies

Method

- Develop a flexible multi-state model
- Age, period or cohort effect
- Informative censoring

Sweden

- Hormone replacement therapy
- Digital mammography
- Different evaluation methods

Collaboration

- Norway
- Finland
- …

THANKS FOR YOUR ATTENTION

- Acknowledgement
  - Umeå University
    - Håkan Jonsson
    - Lennarth Nyström
  - Stockholm-Gotland Regional Cancer Centre
    - Sven Törnberg
    - Klara Miriam Elfström
    - Anna Stoltenberg

Swedish Research Council

NORDIC CANCER UNION

Cancerforsknings-fonden i Norrland

Contact information:
wendy.wu@umu.se