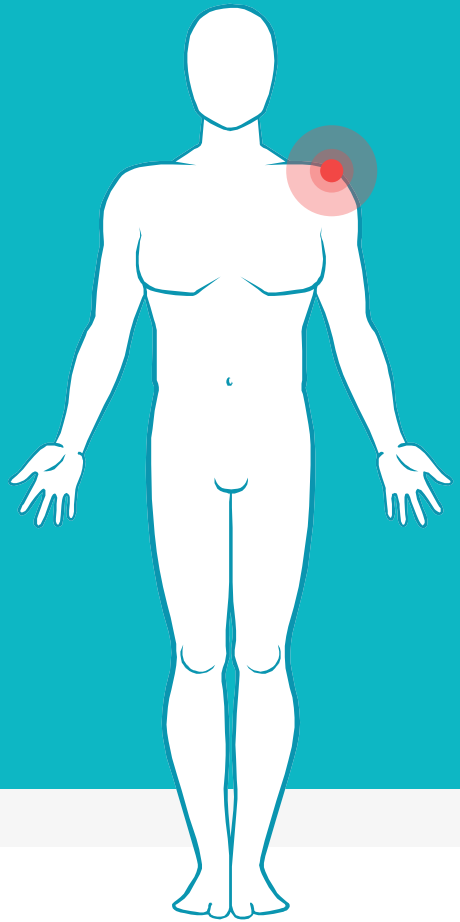
Estimation For Cancer Screening Model

Jiahui Zeng
Mentor : Antonio Olivas



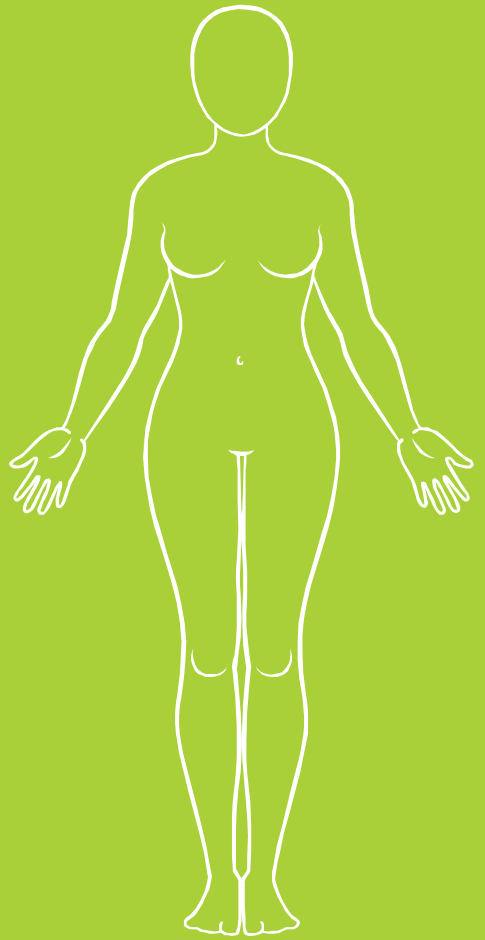
Cancer is accounting for nearly 10,000,000 deaths in 2020, or nearly one in six deaths.

”

2,300,000
Women diagnosed in 2020

685,000
Deaths globally

40%
5 years survival rate in South Africa



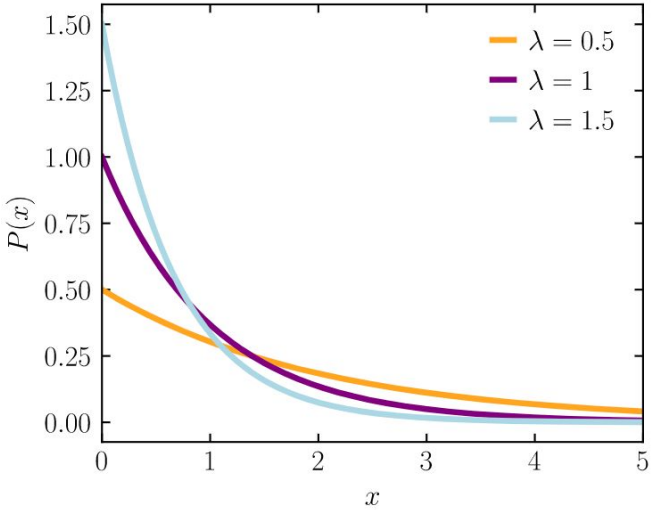
How to reduce breast cancer mortality rate?

Early detection + treatment!

Cancer Screening Program

Three Questions to answer:

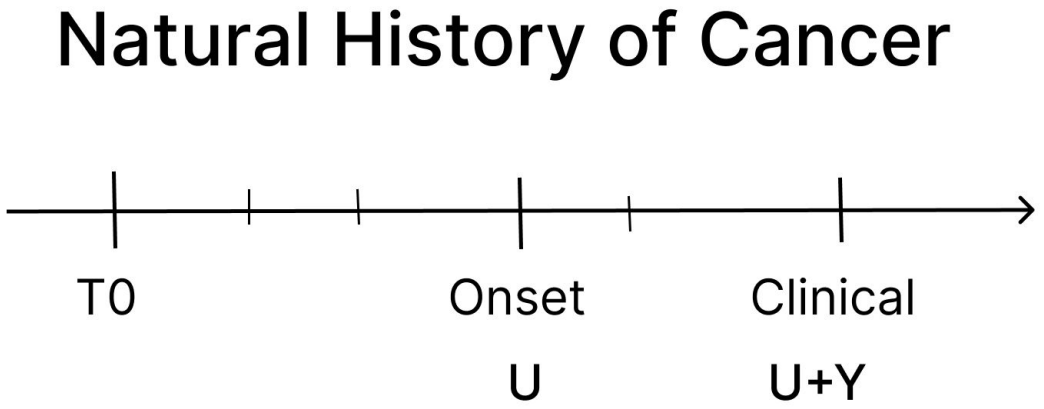
- When to start cancer screening?
  - U ~ Exp(λ)
- How often should the cancer screening be?
  - Y ~ Exp(γ)
- How accurate is the screening tool?
  - Sensitivity(β)



When and How Often to perform cancer screening?

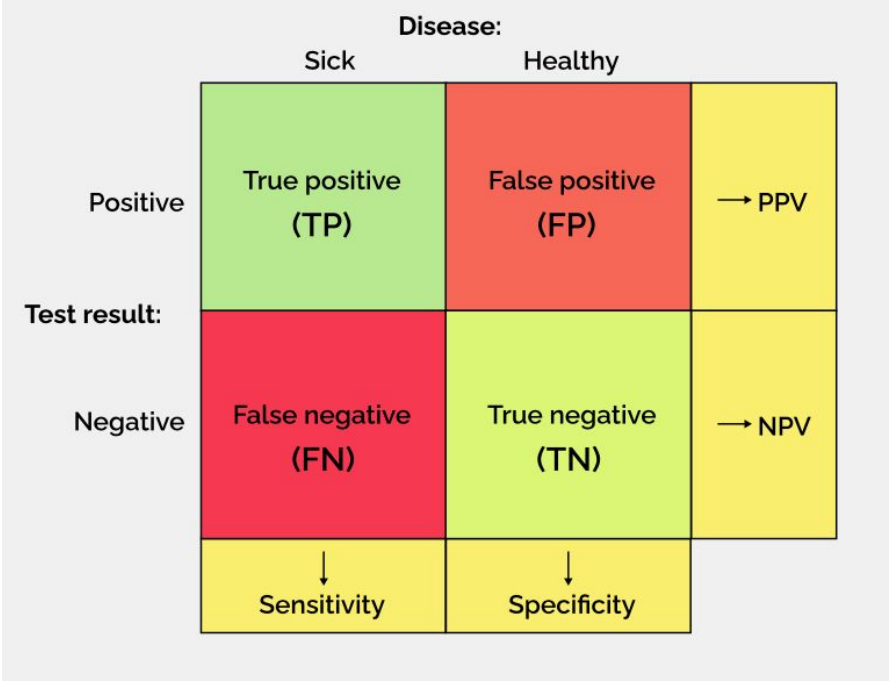
- **T0**: Minimum age to have cancer
- **Onset**: cancer appears in the body without symptoms
- **Clinical**: Symptoms emerge

- **U:** Age of onset ~ Exp(λ) + T0
- **Y**: Sojourn time ~ Exp(γ)
- **U+Y**: Age of clinical

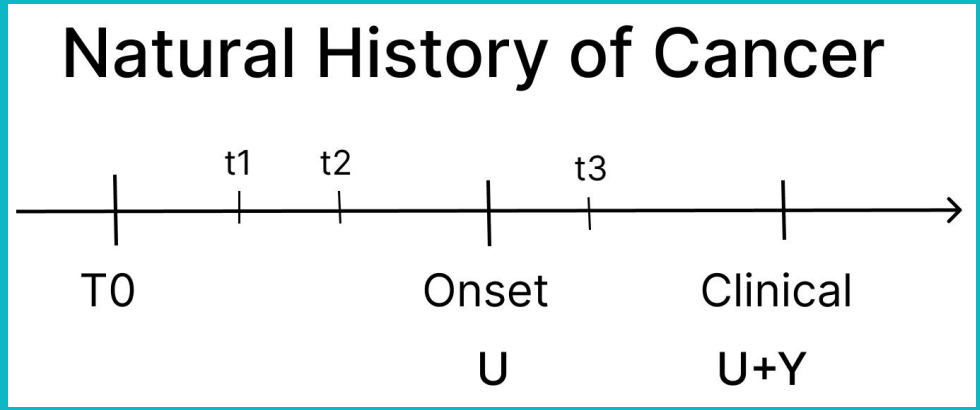


Accuracy of Screening Tool

- **Sensitivity:** β
  - Probability of detecting disease when there is disease
  - **1 - β**: failing to detect disease when there is disease
- **Specificity:** 100%
  - No False Positive



Estimating the Parameters: λ, γ, β



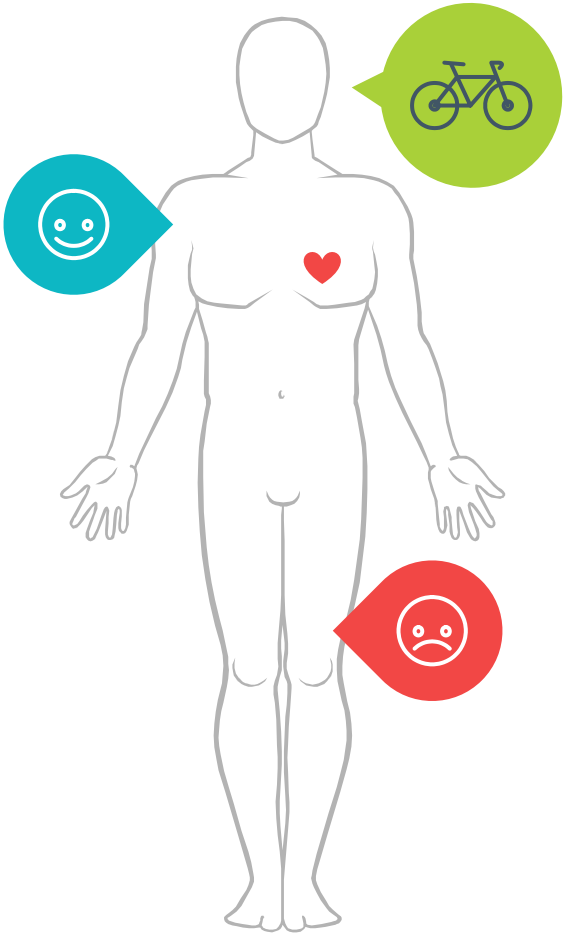
What do we hope to gain?

Recall: the expected value E[X] of an exponential distribution is 1 over the parameter

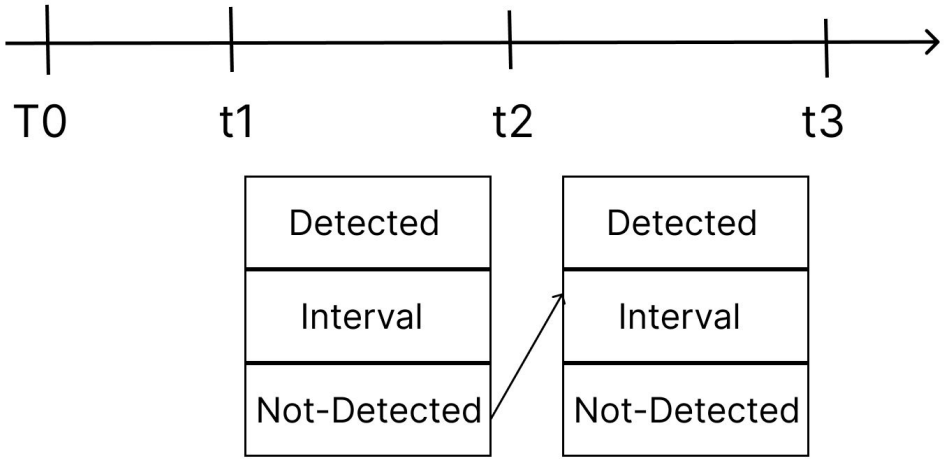
- Mean sojourn time

Maximum Likelihood Estimation

A method that determines the parameter of a function



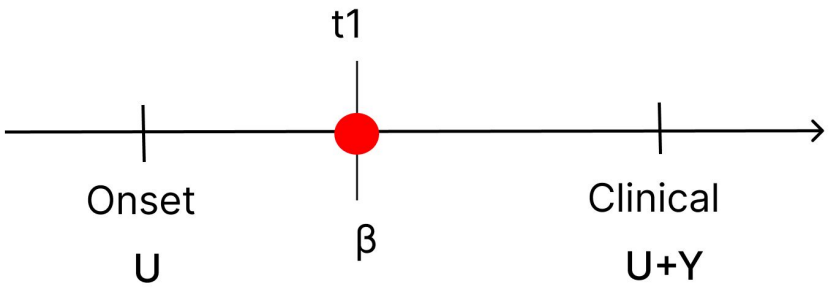
Model



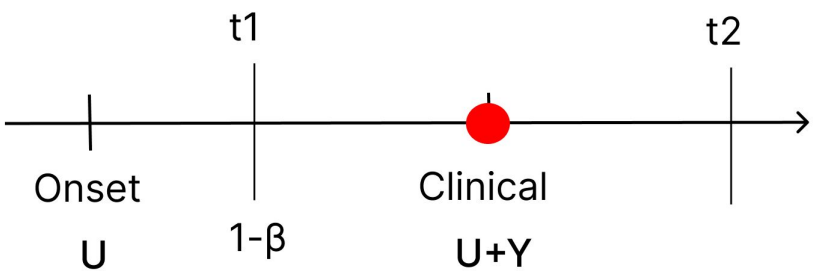
- Multinomial
  - Screen-detected cases
  - Interval-detected cases
    - Rapid development between screenings
  - Not-detected cases

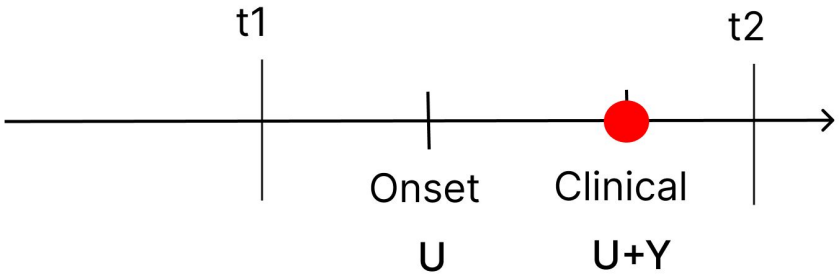
$t1 \leq a < t2$

Screen-Detected Case



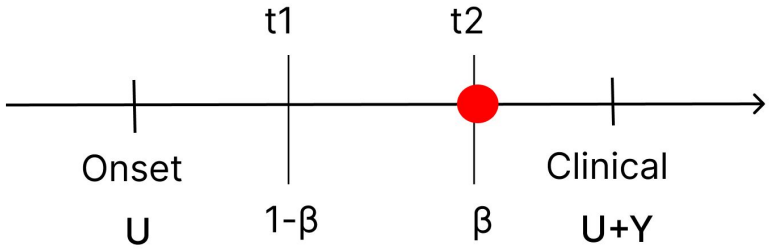
Interval-Detected Case

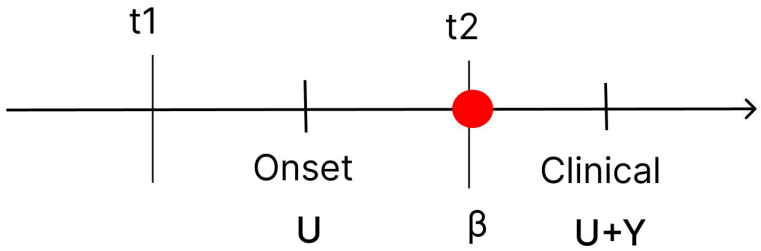




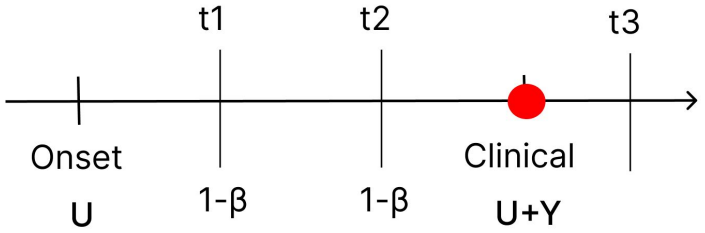
$t2 \leq a < t3$

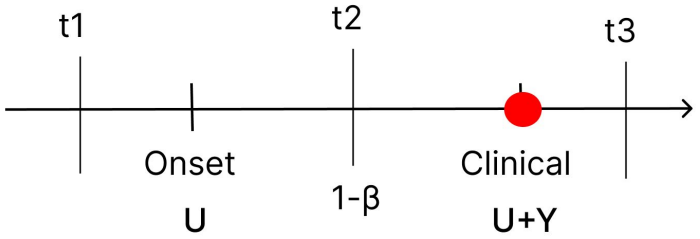
Screen-Detected Case

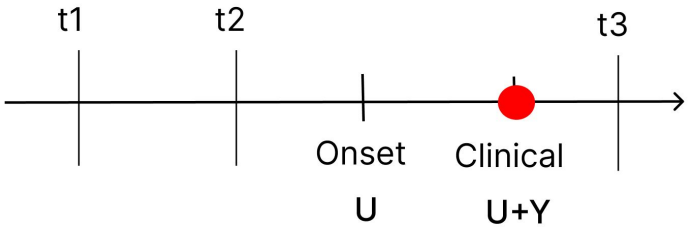




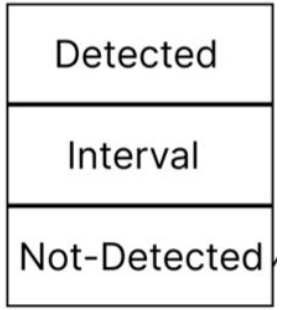
Interval-Detected Case







Model



- S_i = screen detected at t_i
- I_i = interval cases between t_i and t_{i+1}
- s_i = number of screen-detected at t_i
- i_i =number of interval cases between t_i and t_{i+1}
- n_i = total number of screened at t_i

- $L(\lambda, \gamma, \beta) = P(S_1)^{s1} * P(I_1)^{i1} * (1 - S_1 - I_1)^{n1 - s1 - i1} * P(S_2)^{s2} * P(I_2)^{i2} * (1 - S_2 - I_2)^{n2 - s2 - i2}$

- $\ell(\lambda, \gamma, \beta) = s_1 \ln(P(S_1)) + i_1 \ln(P(I_1)) + (n_1 - s_1 - i_1)\ln((1 - S_1 - I_1)) +$

$$s_2 \ln(P(S_2)) + i_2 \ln(P(I_2)) + (n_2 - s_2 - i_2)\ln((1 - S_2 - I_2))$$

Application to Real Data

Web Table 1

Screening round	No. of women	Screen-detected cases	Interval-detected cases
1	19711	142	15
2	17669	66	10
3	17347	43	9
4	17193	54	9
5	9876	28	5

CNBSS-2. Grouped data from the Canadian Breast Cancer Screening Study-2 [3]. "No. of women" is the number of women who attended all screening rounds up to and including the current round.

Result

```{r}
data <- rbind(c(142, 15, 19711-142-15), c(66, 10, 17669-66-10))
known <- data.frame(t0 = 45, t1 = 50, t2 = 51, t3 = 52)
known <- cbind(known,
               s1 = data[1, 1],
               r1 = data[1, 2],
               n1 = data[1, 3],
               s2 = data[2, 1],
               r2 = data[2, 2],
               n2 = data[2, 3])
esti <- maxLik(logLik = max_likelihood, start = c(0.001, 0.01, 0.6),
       method = "NR", input = known)

coef(esti)
```

```
[1] 0.00317097 0.19592243 0.72099802
```

Result

Web Table 2

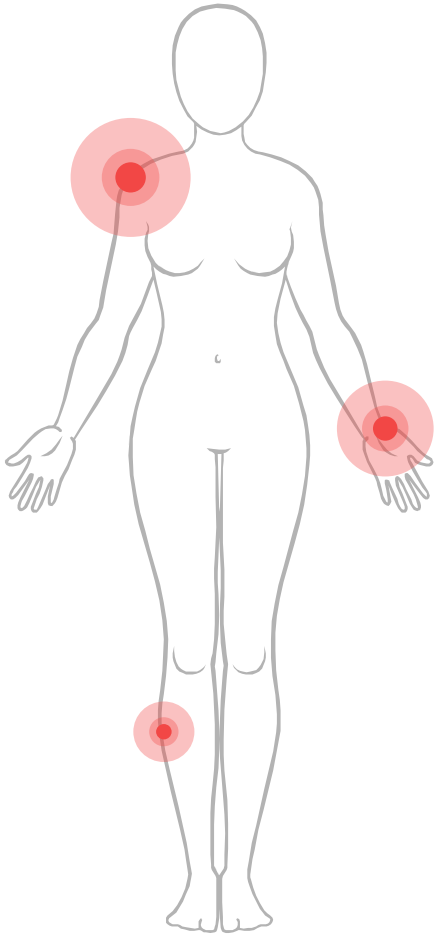
ψ	$\hat{\lambda}$	95% CI	$\hat{w}$	95% CI	$\hat{\beta}$	95% CI	χ^2	P-value
0.00	0.30	0.10-0.47	0.0031	0.0023-0.0036	0.81	0.44- NA	6.20	0.7984
0.05	0.30	0.09-0.48	0.0029	0.0023-0.0034	0.75	0.41-0.96	7.03	0.7226
0.10	0.28	0.09-0.48	0.0028	0.0023-0.0032	0.69	0.40-0.90	7.74	0.6546
0.20	0.26	0.11-0.48	0.0026	0.0022-0.0030	0.59	0.39-0.79	8.23	0.6061
0.40	0.34	0.16-0.62	0.0024	0.0022-0.0027	0.51	0.38-0.64	8.39	0.5905

CNBSS-2: Parameter estimation and goodness of fit for constrained mixture model. The mixture model with fixed fraction of indolent cancers ψ is fit to the CNBSS-2 data (see Web Table 1). Onset of preclinical disease is assumed to be negligible before age $\Delta_0 = 45$ years.

- $\lambda \approx 0.00317$
  - 1% women have breast cancer at 48.2 years old
- $\gamma \approx 0.1959$
  - Average sojourn time = 5 years
- $\beta \approx 0.721$

Next Steps

- Further evaluate our result to produce a meaningful conclusion.
- Perform simulations to validate our model
- Consider other possible distributions



Thank you for listening!

Special thanks to my mentor Antonio and the DRP!

Reference

"Breast Cancer." *World Health Organization*, World Health Organization, https://www.who.int/news-room/fact-sheets/detail/breast-cancer

"Cancer." *World Health Organization*, World Health Organization, https://www.who.int/news-room/fact-sheets/detail/cancer

Ryser MD, Gulati R, Eisenberg MC, Shen Y, Hwang ES, Etzioni RB. Identification of the Fraction of Indolent Tumors and Associated Overdiagnosis in Breast Cancer Screening Trials. Am J Epidemiol. 2019 Jan 1;188(1):197-205. doi: 10.1093/aje/kwy214. PMID: 30325415; PMCID: PMC6321806.

https://www.omnicalculator.com/statistics/sensitivity-and-specificity

CREDITS

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash