An Introduction to Longitudinal Biomarker Evaluations

Margaret Pepe and Tracey Marsh

Public Health Sciences Division
Fred Hutchinson Cancer Research Center, Seattle, WA

EDRN — DMCC

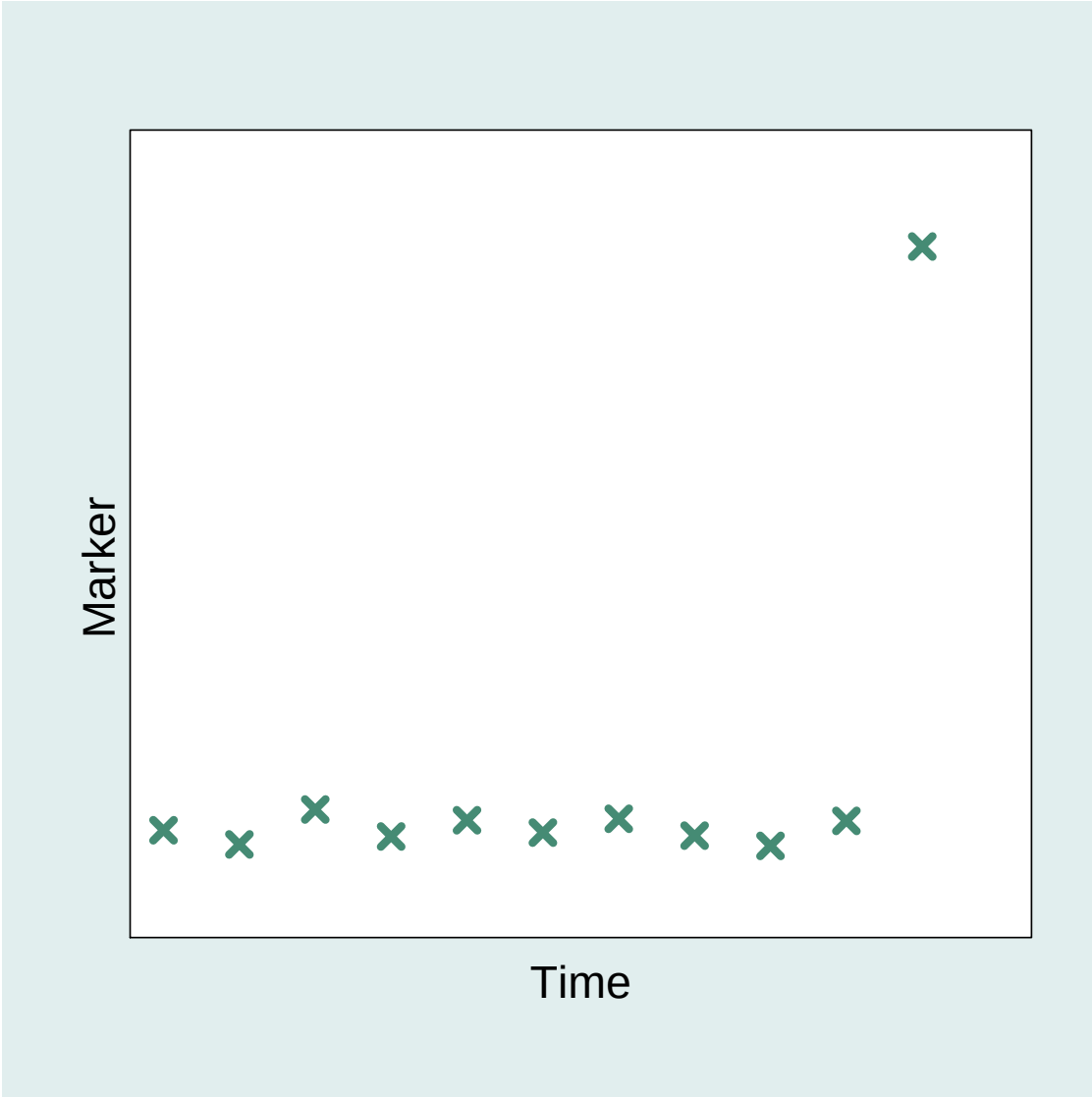
FRED HUTCH
CURES START HERE®

What is longitudinal biomarker data?

- Biomarker measured in serial samples from each individual.
- CA-125 measured annually for ovarian cancer screening.
- CEA measured monthly for colon cancer recurrence after initial treatment.

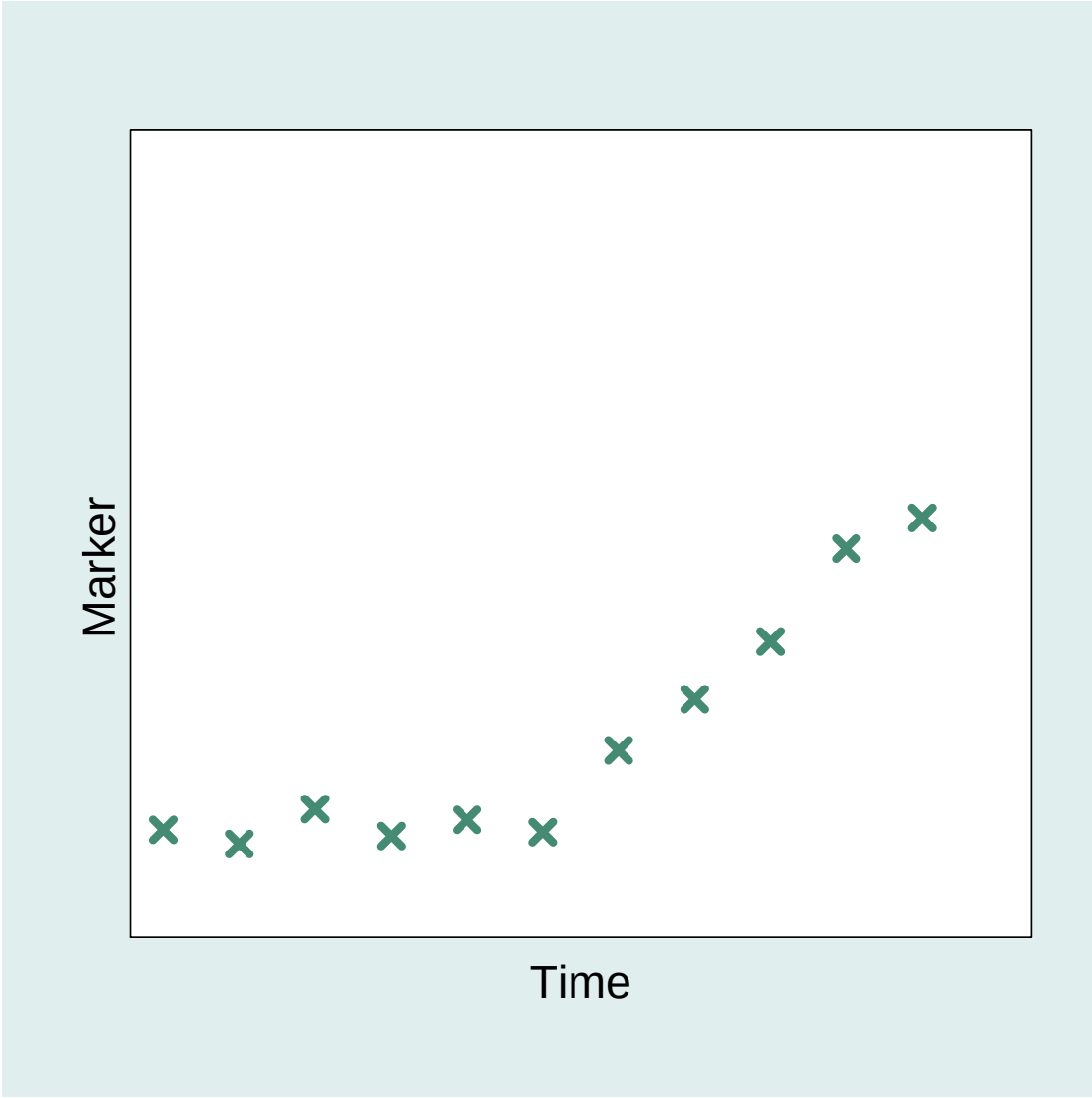
Key Idea: Change in biomarker over time may inform about disease onset.

- **Example**: a large change from relatively stable levels.



Key Idea: Change in biomarker over time may inform about disease onset.

- **Example**: consistent trend away from baseline.



Not talking about markers

- with undetectable levels in controls
- germline gene variants or other lifelong stable measures of risk

Talking about markers

- present in people without cancer (controls)
- controls tend to have their own individual set-point

Talking about

- CA-125 for ovarian cancer screening (Skates).
- PSA for prostate cancer screening (Zheng).
- AFP, DCP for liver cancer screening (Tayob).

Intra-class Correlation Coefficient (ICC)

$$\text{ICC} = \frac{\text{variation between subjects}}{\text{var(between)} + \text{var(within)}}$$

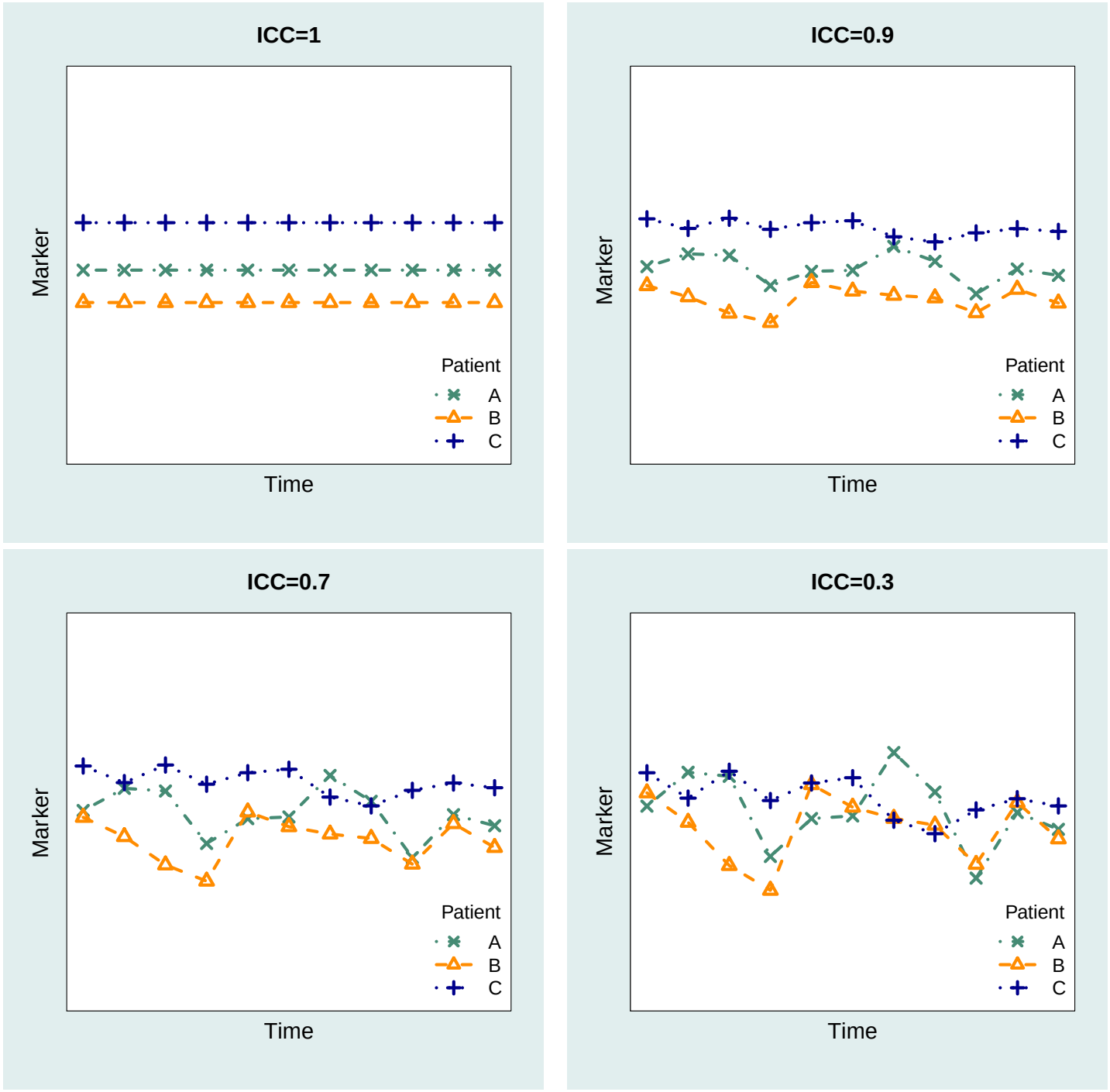
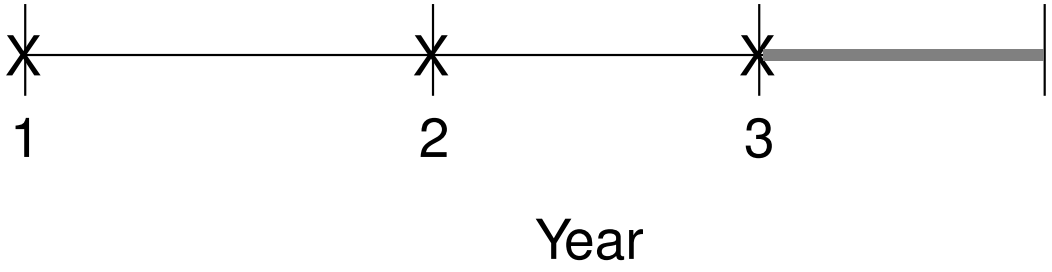
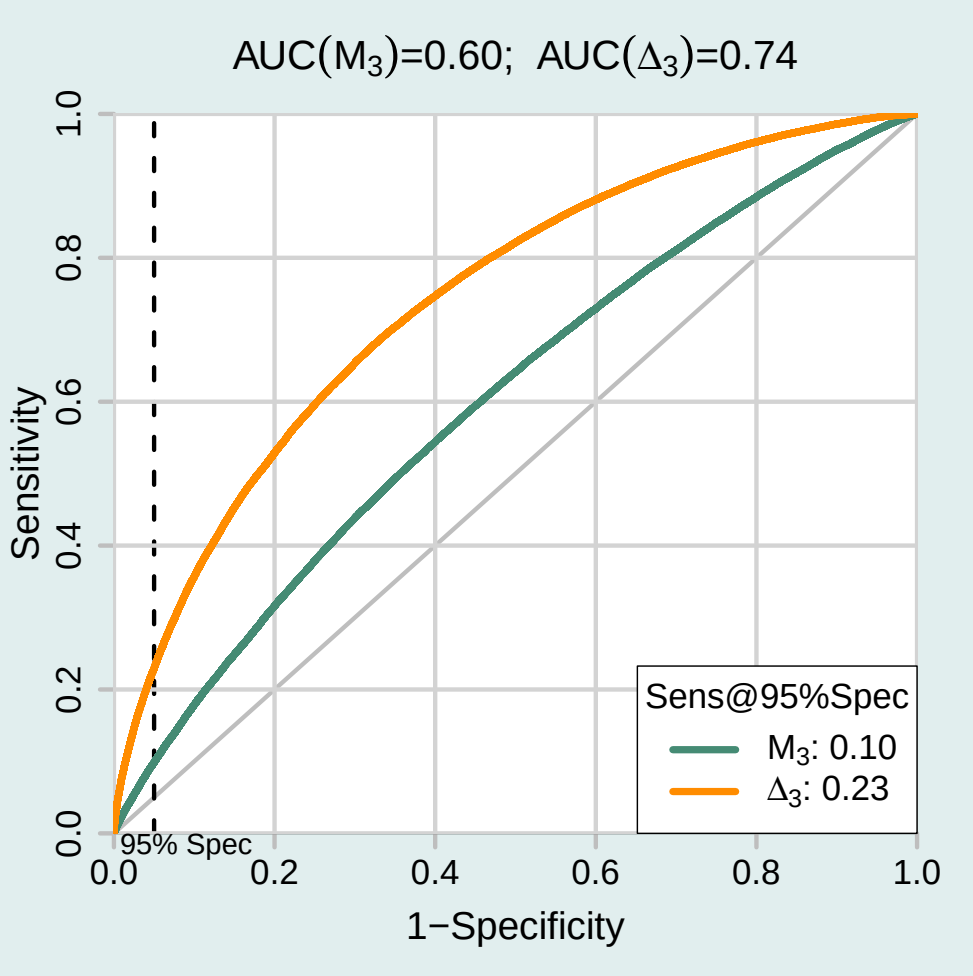

ICC=1
Marker
Time
Patient
A
B
C
ICC=0.9
Marker
Time
Patient
A
B
C
ICC=0.7
Marker
Time
Patient
A
B
C
ICC=0.3
Marker
Time
Patient
A
B
C

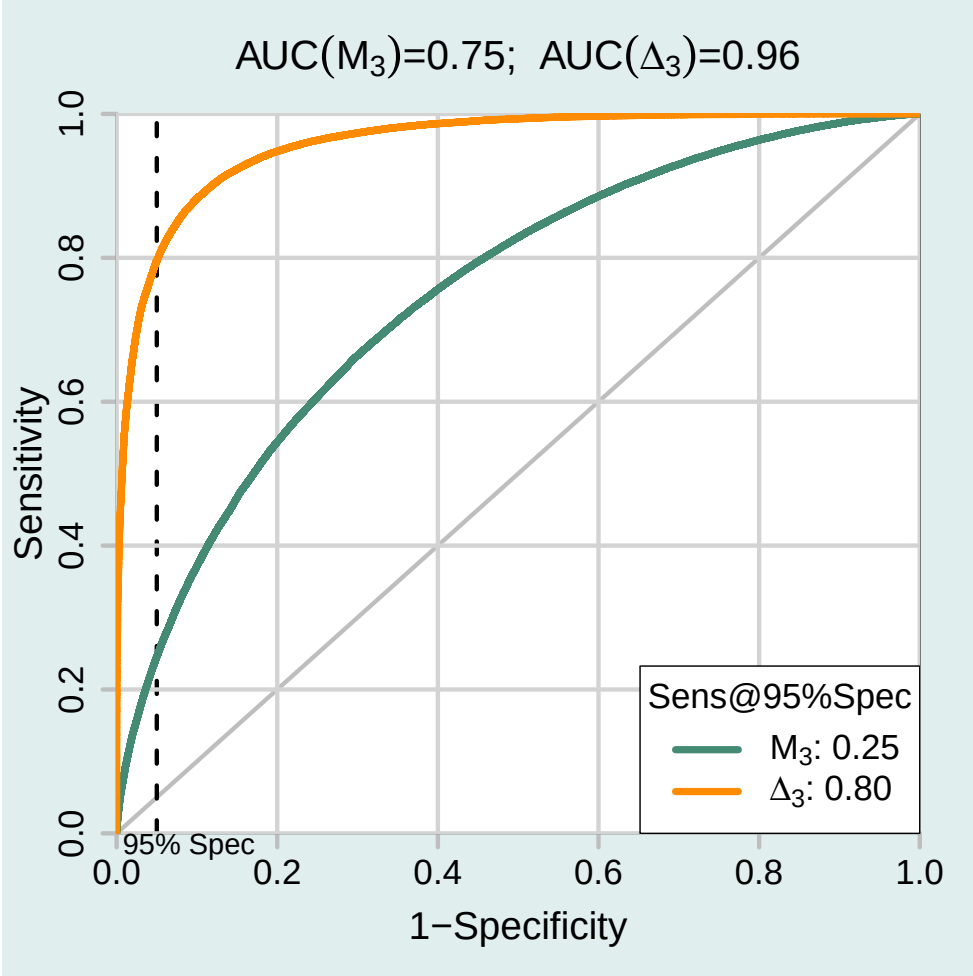
Illustration of potential added value from longitudinal biomarker

- Stored annual samples
- Could we have detected cases that occurred after start of year 3 using markers measured up to year 3?
- Single time point marker: M_3
- Longitudinal marker: $\Delta_3 = M_3 - ave(M_1, M_2)$

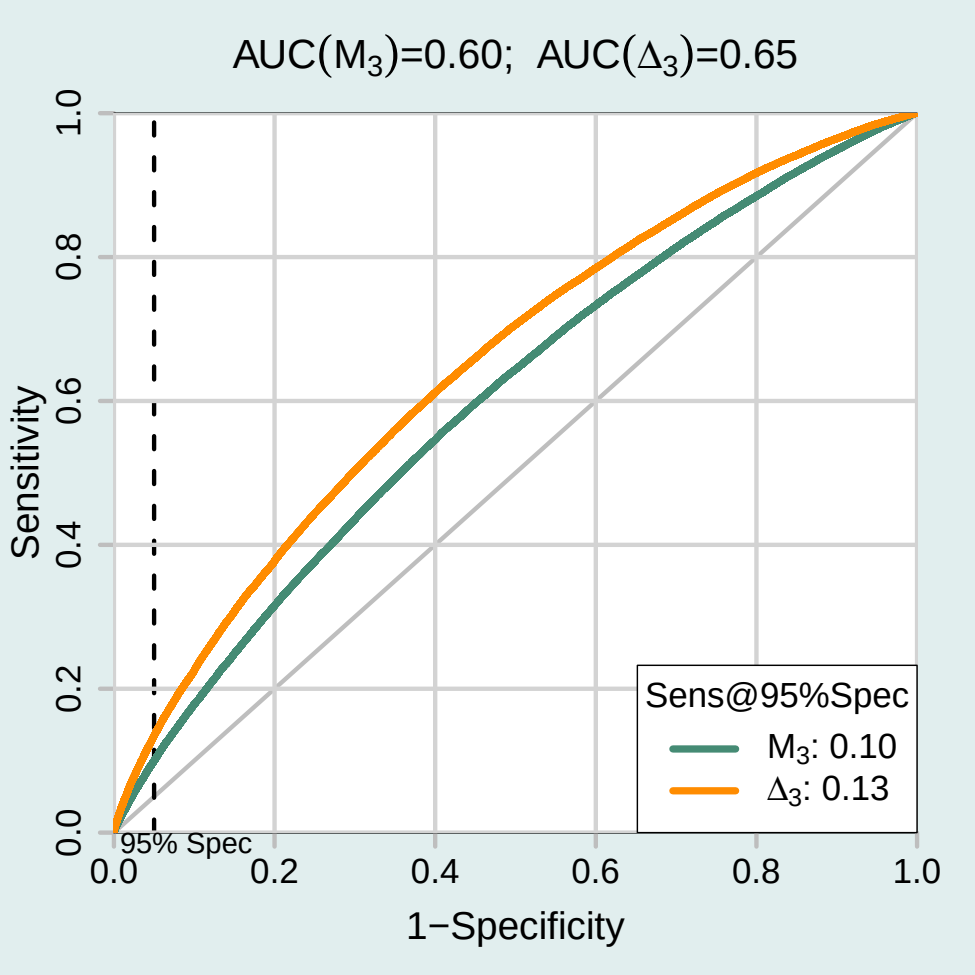


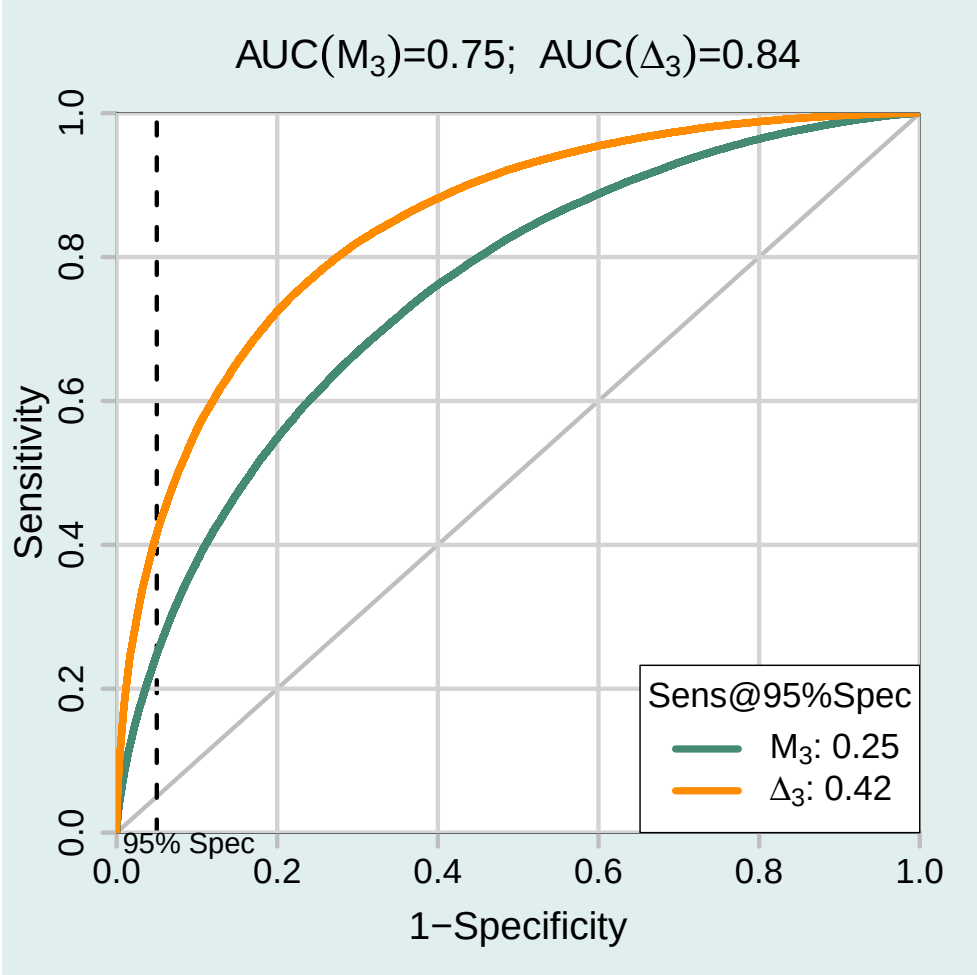
Low within-subject variation in controls (ICC=0.90)



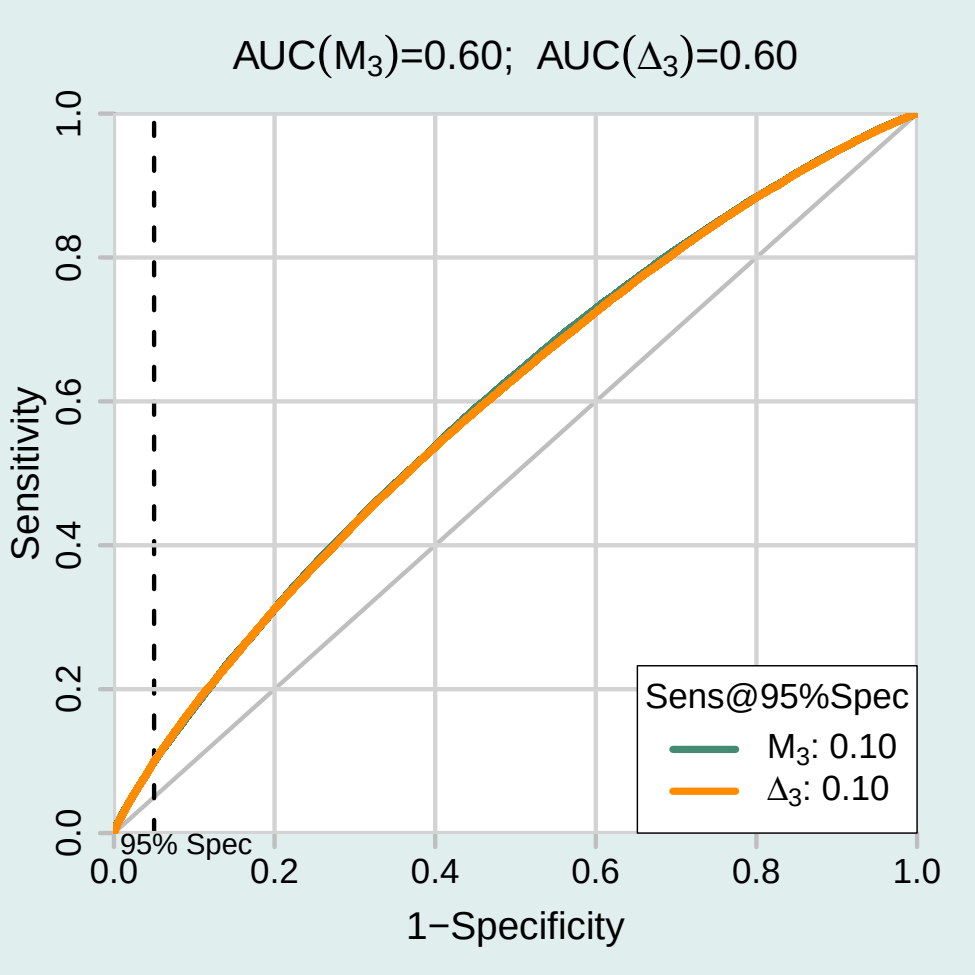


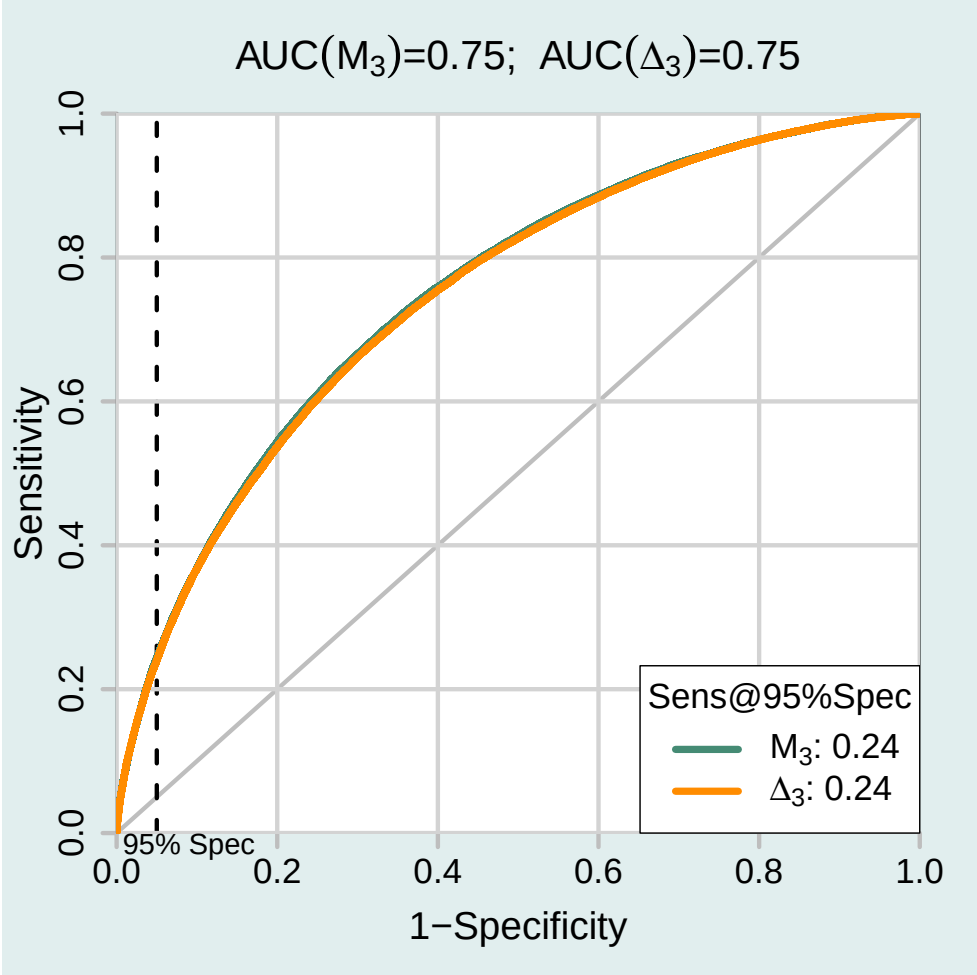
Moderate within-subject variation (ICC=0.70)





High within-subject variation (ICC=0.30)





Recommendations

- Markers that show mild/moderate performance at a single time point
- Consider if there is potential for better performance with longitudinal ascertainment.
- Statistical considerations.
  - Do a preliminary study using serial samples *from controls* to assess ICC
  - This is cheap compared with larger study to assess biomarker performance for detecting disease in cases.
  - Assay variability.
- Feasibility considerations and complications
  - Will people return to provide serial samples?
  - Algorithms to accommodate missing time points.

Algorithms for using longitudinal biomarker data with real data in real applications

- Ovarian cancer — Steven Skates
- Prostate cancer — Yingye Zheng
- Liver cancer — Mabihah Tayob

Appendix

$$\begin{aligned} & \text{AUC}(M_3) = 0.60 = \Phi(a/\sqrt{2}) \\ \Rightarrow \quad & \text{ROC}(t) \quad = \Phi(a + b\Phi^{-1}(t)) \end{aligned}$$

To generate data

$$M_{ij} = I[i = 1]D_j + \mu_j + \varepsilon_{ij}$$

i = time $\quad$ j = subject

$$M_{ij} = I[i = 3]D_j + \mu_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} = N(0, \sigma^2_\varepsilon) \quad \cdot \quad \sigma^2_\varepsilon = 1$$

$$\cdot \mu_j = N(0, \sigma^2_\mu) \qquad \sigma^2_\mu = \frac{\text{ICC}}{1 - \text{Icc}}$$

$$D_j = I_M \times \mu_D$$

$$I_M = \text{binary} \begin{cases} \text{prob} = 100\% & \text{marker in 100\% cases} \\ = 50\% & \text{marker in 50\% cases} \end{cases}$$

$$\mu_D = a \times \sqrt{(\sigma^2_\mu + \sigma^2_\varepsilon)}$$