ROC-based methods for meta-analysis of prognosis studies

Satoshi Hattori

Department of Biomedical Statistics, Osaka University, Japan

Collaborators:

Xiao-Hua Zhou, Yi Zhou (Peking U), Ao Huang (Goettingen U)

"Recent Advances in Meta-Analysis: Methods and Software" Symposium, 23Aug2023, UMG Goettingen

Outline

To outline our recent developments of time-dependent Summary ROC-based methods for meta-analysis (MA) of prognosis studies with a survival outcome

1. Motivating examples
2. Brief review of summary ROC (SROC) for diagnosis(binary) MA
   - Bivariate Gaussian model
   - Bivariate binomial model
3. Time-dependent SROC for prognosis MA
4. Assessing potential impacts of publication bias (PB)
5. Concluding remarks

Motivating examples
and
Brief review of SROC for diagnosis MA

Prognosis studies

- To evaluate usefulness of factors (including biomarkers) in discriminating who will or will not occur an event of interest in future.

- Useful in current clinical practice
  - To understand disease.
  - To seek potential targets to improve future outcome in patients.

Prognosis studies for a continuous biomarker

- We consider MA of prognosis studies for a continuous biomarker with a survival outcome.

- In general, prognosis studies handle any types of outcomes (continuous, binary and survival).

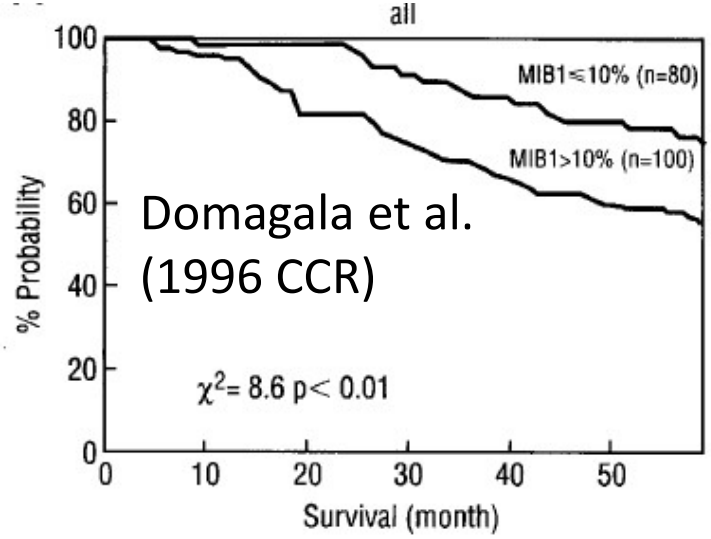
- In this talk, we call a biomarker study with

  a binary outcome *Diagnosis* study

  a survival outcome *Prognosis* study

Heterogeneous cut-off value issue in diagnosis/prognosis MA

- Diagnosis/prognosis studies often employ a study-specific (heterogeneous) cut-off value and define high-/low-expression groups
  - Sensitivity/specificity, diagnosis odds ratio (OR) in diagnosis studies
  - Kaplan-Meier estimates of the two groups and hazard ratio (HR)



- Simple aggregation of ORs/HRs with the standard MA is hard to interpret.
- Summary ROC is an appealing tool in diagnosis MA(binary)

Example of diagnosis MA: Troponin data Becattini et al. (Circulation 2007)

- Meta analysis of 20 prognostic studies (1985 patients)
- Troponin I/T in acute pulmonary embolism
- Endpoint: short-term death (binary)
- Two troponins have different scales.
- Cut-off values vary across studies:
  - Troponin I: 0.06 - 2.0, Troponin T: 0.01 – 0.1

- Combined OR (Mixed-effect model):
  - Troponin I: OR=4.01 (95% CI: 2.23 – 7.23)
  - TroponinT: OR=7.95 (95% CI: 3.79 – 16.65)
- Hard to interpret and compare these ORs.

Summary receiver operating characteristics (SROC) for diagnosis MA: data structure

S: No of published studies

D: binary outcome

M: biomarker

$c^{(s)}$: cut-off value

X=I$\left(c^{(s)} \leq M < \infty\right)$:high/low-expression

The following frequencies are supposed to be observed for the sth study (s = 1,2,… S)

	$D=0$	$D=1$	
$X=0: 0 \leq M < c^{(s)}$	$N_{00}^{(s)}$	$N_{01}^{(s)}$	$n_{0+}^{(s)}$
$X=1: c^{(s)} \leq M < \infty$	$N_{10}^{(s)}$	$N_{11}^{(s)}$	$n_{1+}^{(s)}$
	$n_{+0}^{(s)}$	$n_{+1}^{(s)}$	$n^{(s)}$

SROC for diagnosis MA: idea

- Due to heterogeneous cut-off values over studies, sensitivity and specificity are likely to be negatively correlated.

- Model the association between sensitivity and specificity

with bivariate mixed-effect models

  - Bivariate Gaussian model
  - Bivariate binomial model

- Express sensitivity as a function of specificity (SROC curve)

Bivariate Gaussian model for diagnosis MA

Model the joint distribution of empirical sensitivity and specificity based on asymptotic normal approximation

$$\hat{\mu}_{sen}^{(s)} = \text{logit}\left(\hat{sen}^{(s)}\right) = \text{logit}\left(\frac{N_{11}^{(s)}}{n_{+1}^{(s)}}\right)$$

$$\hat{\mu}_{spe}^{(s)} = \text{logit}\left(\hat{spe}^{(s)}\right) = \text{logit}\left(\frac{N_{00}^{(s)}}{n_{+0}^{(s)}}\right)$$

Diagonal matrix
Easily estimated

$$\begin{pmatrix}\hat{\mu}_{sen}^{(s)}\\ \hat{\mu}_{spe}^{(s)}\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_{sen}^{(s)}\\ \mu_{spe}^{(s)}\end{pmatrix}, U^{(s)}\right), \qquad U^{(s)}(t) = \begin{pmatrix}\hat{s}_{sen}^{(s)}, 0\\ 0, \hat{s}_{spe}^{(s)}\end{pmatrix}$$

$$\begin{pmatrix}\mu_{sen}^{(s)}\\ \mu_{spe}^{(s)}\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_{sen}\\ \mu_{spe}\end{pmatrix}, \Sigma\right), \qquad \Sigma = \begin{pmatrix}\tau_{sen}^2, & \tau_{sen,spe}\\ \tau_{sen,spe}, & \tau_{spe}^2\end{pmatrix}$$

ML or REML estimation is applicable

$$SROC\ curve: E\left(\mu_{sen}^{(s)} | \mu_{spe}^{(s)}\right)$$

Reitsma et al. (2005, J Clinical Epidemiology)

Bivariate binomial model for diagnosis MA

Model the joint distribution of the bivariate binomial random variables.

$$\pi_1^{(s)} = TPR^{(s)} = s\hat{e}n^{(s)} = P(X = 1|D = 1)$$
$$\pi_0^{(s)} = FPR^{(s)} = 1 - s\hat{p}e^{(s)} = P(X = 1|D = 0)$$

$$N_{1d}^{(s)} \sim Bin\left(n_{+d}^{(s)}, \pi_d^{(s)}\right), \qquad \text{logit}\left(\pi_d^{(s)}\right) = \frac{\theta + \theta^{(s)} + \left(\alpha + \alpha^{(s)}\right)Z_d^{(s)}}{\exp\left(\beta Z_d^{(s)}\right)}, \quad \text{d=0,1}$$

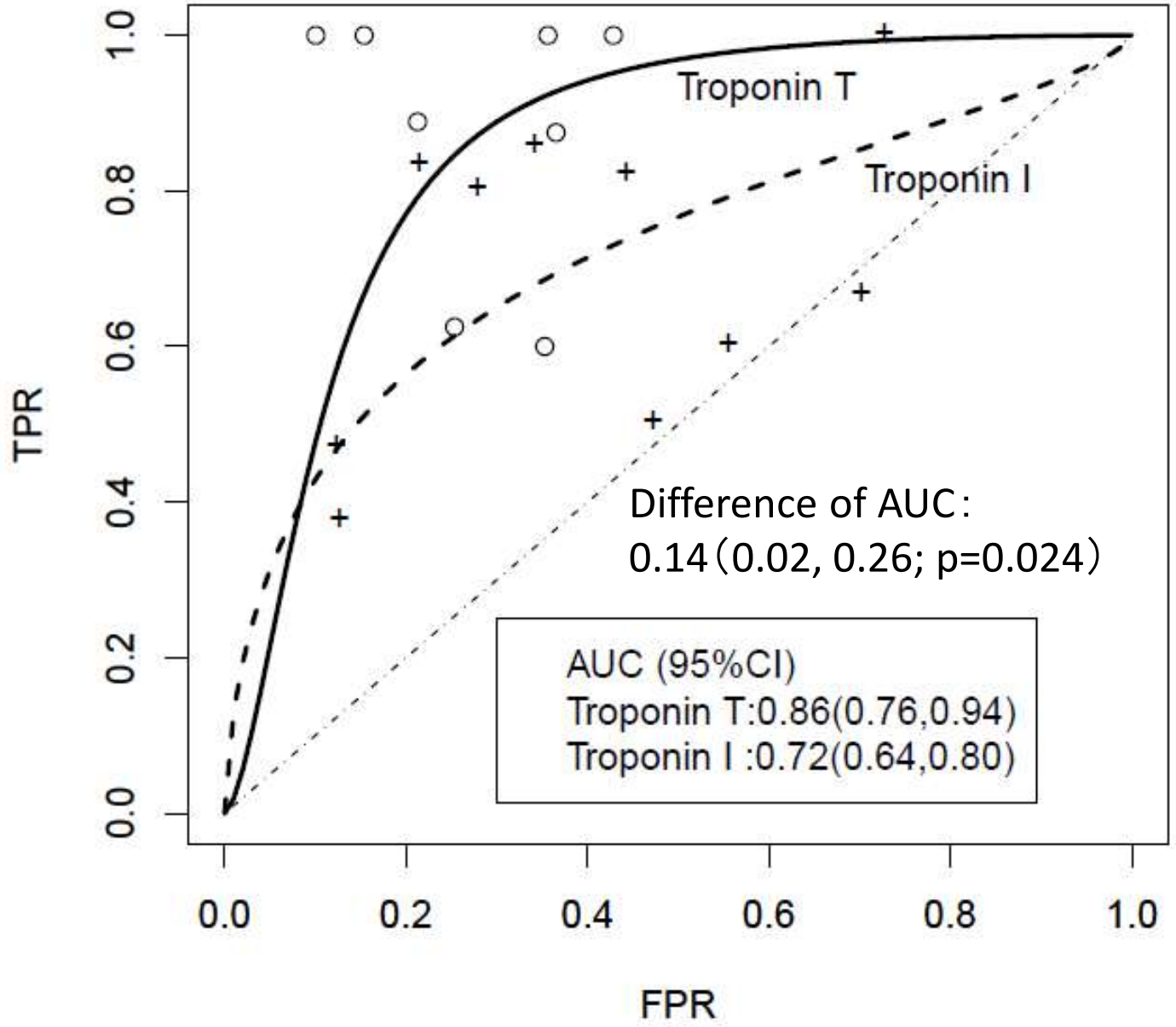
$$\mathrm{Z}_d^{(s)} = -1/2(d = 0), \qquad = 1/2(d = 1)$$
$$\theta^{(s)} \sim N\left(0, \sigma_\theta^2\right), \alpha^{(s)} \sim N(0, \sigma_\alpha^2), \theta^{(s)} \perp \alpha^{(s)}$$

MCMC（Rutter and Gastnis 2001, SiM）
REML（Macaskill 2004, J. Clinical Epidemiology）

SROC curve：

$$TPR = \frac{1}{1 + \exp\{-(\alpha \exp(-0.5\beta) + \text{logit}(FPR)\exp(-\beta))\}}$$

SROC with bivariate binomial model: Troponin data



Hattori and Zhou (2016, SiM)

Time-dependent SROC for prognosis MA

Example of prognosis MA: Ki-67 data
De Azambuja et al. (British J Cancer 2007)

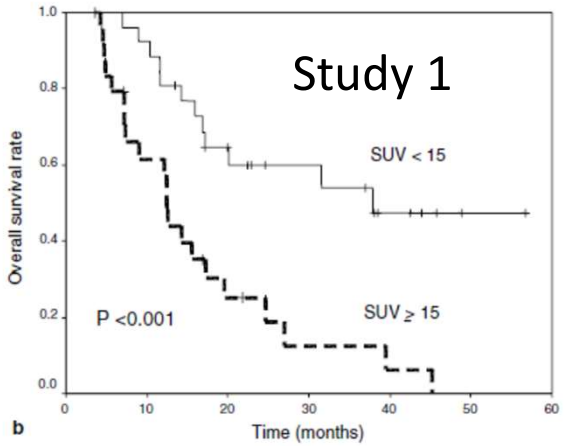
- Meta-analysis of 38 prognostic studies with 9472 subjects.
- Ki-67 in early breast cancer
  - Proportion of expressed cells in tumor (0-1)
  - Proliferation
  - Important role in discriminating between Luminal A and B
- Endpoint: OS

- Standard meta-analysis techniques were applied:
  - Random-effect model: HR=1.93 (95%CI 1.74-2.14)
- Cut-off values are heterogeneous (0.035 – 0.32)

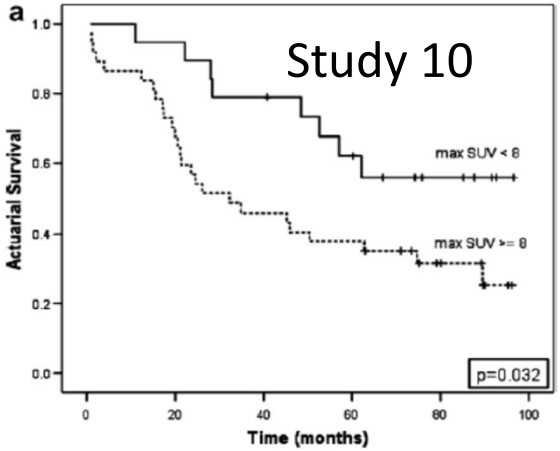
  ⇒The above results seem to be difficult to interpret.

Time-dependent Summary ROC curve

Data: Kaplan-Meier plot



Heterogeneous cut-off values



MA version of time-dependent ROC (Heagerty et al. 2000, Biometrics)

Combescure et al. (2016, SMMR)
-Spline-based joint model to KM and biomarker dist.

Time-dependent (summary) ROC curve

Hattori and Zhou (2016, SiM)
-Natural extension of sROC for diagnostic studies (binary outcome)
-Bivariate Gaussian model
-Bivariate binomial model

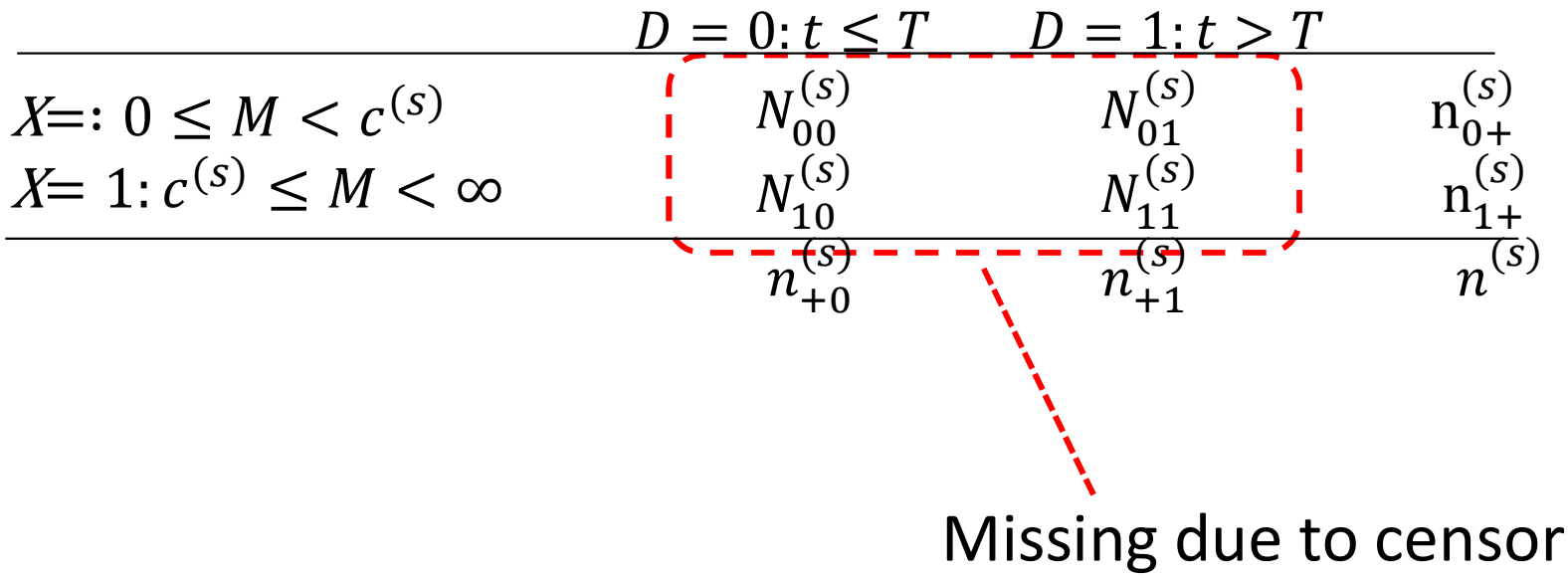
Time-dependent SROC for prognosis MA

Suppose we are interested in estimating SROC at t-year.

Evaluate capacity of the biomarker M to “diagnose” t-year survivor

T: survival time

$D=I(T \leq t)$: unobservable due to censoring



Bivariate Gaussian model for prognosis MA

Idea: construct time-dependent sensitivity and specificity with KMs.

$$sen^{(s)}(t) = P\left(M \geq c^{(s)} | T \leq t\right) = \frac{\left\{1 - S_1^{(s)}(t)\right\} q_1^{(s)}}{\left\{1 - S_1^{(s)}(t)\right\} q_1^{(s)} + \left\{1 - S_0^{(s)}(t)\right\} q_0^{(s)}}$$

$$spe^{(s)}(t) = P\left(M < c^{(s)} | T > t\right) = \frac{S_0^{(s)}(t) q_0^{(s)}}{S_1^{(s)}(t) q_1^{(s)} + S_0^{(s)}(t) q_0^{(s)}}$$

$$q_1^{(s)} = P\left(M \geq c^{(s)}\right), q_0^{(s)} = P\left(M < c^{(s)}\right)$$

Consider empirical versions: $\hat{sen}^{(s)}(t), \hat{spe}^{(s)}$

Bivariate Gaussian model for prognosis MA

No longer diagonal
But it is estimable with KMs

$$\hat{\mu}_{sen}^{(s)}(t) = \text{logit}\left(\hat{sen}\left(v^{(s)}, t\right)\right)$$
$$\hat{\mu}_{spe}^{(s)}(t) = \text{logit}\left(\hat{spe}\left(v^{(s)}, t\right)\right)$$

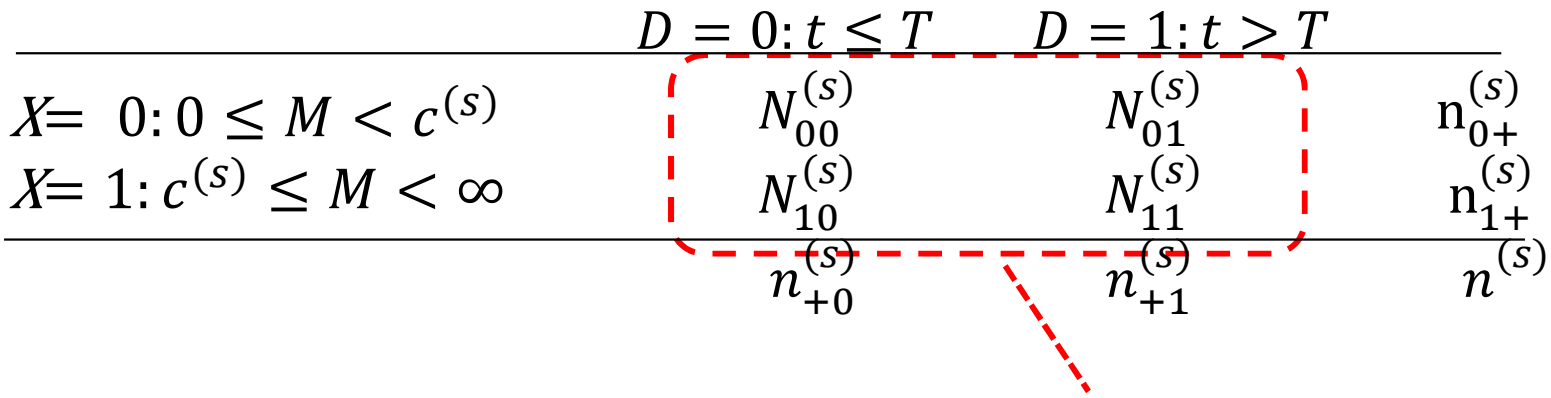
$$\begin{pmatrix} \hat{\mu}_{sen}^{(s)}(t) \\ \hat{\mu}_{spe}^{(s)}(t) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{sen}^{(s)}(t) \\ \mu_{spe}^{(s)} \end{pmatrix}(t), U^{(s)}(t)\right), \quad U^{(s)}(t) = \begin{pmatrix} \hat{s}_{sen}^{(s)}(t), & \hat{s}_{sen,spe}^{(s)}(t) \\ \hat{s}_{sen,spe}^{(s)}(t), & \hat{s}_{spe}^{(s)}(t) \end{pmatrix}$$

$$\begin{pmatrix} \mu_{sen}^{(s)}(t) \\ \mu_{spe}^{(s)}(t) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{sen}(t) \\ \mu_{spe}(t) \end{pmatrix}, \Sigma(t)\right), \qquad \Sigma(t) = \begin{pmatrix} \tau_{sen}^2(t), & \tau_{sen,spe}(t) \\ \tau_{sen,spe}(t), & \tau_{spe}^2(t) \end{pmatrix}$$

$$SROC\ curve: E\left(\mu_{sen}^{(s)}(t) | \mu_{spe}^{(s)}(t)\right)$$

Bivariate binomial model for prognosis MA

If missing cell frequencies are successfully imputed, the standard bivariate binomial model for diagnosis MA is applicable.



Problem: how to impute the cell frequencies?

Bivariate binomial model for prognosis MA

Multiple imputation (MI):

Sample missing observations from conditional distribution given observed:

$$\mathrm{Y}_{mis} \sim f(\mathrm{Y}_{mis}|\mathrm{Y}_{obs})$$

Rubin's variance formula is justified for these samples.

Bivariate binomial model for prognosis MA:

	$D=0: t \le T$	$D=1: t > T$	
$Z=0: 0 \le M < c^{(s)}$	$N_{00}^{(s)}$	$N_{01}^{(s)}$	$n_{0+}^{(s)}$
$Z=1: c^{(s)} \le M < \infty$	$N_{10}^{(s)}$	$N_{11}^{(s)}$	$n_{1+}^{(s)}$
	$n_{+0}^{(s)}$	$n_{+1}^{(s)}$	$n^{(s)}$

$$Y_{mis}: \hat{p}_z^{(s)} = \frac{N_{z0}^{(s)}}{n_{z+}^{(s)}}, \qquad Y_{obs}: \hat{S}_z^{(s)}(\mathrm{t})$$

$$f\left(\hat{p}_z^{(s)} | \hat{S}_z(\mathrm{t})\right) = \int_0^1 f\left(\hat{p}_z^{(s)} | \hat{S}_z^{(s)}(\mathrm{t}), S_z^{(s)}(\mathrm{t})\right) f\left(S_z^{(s)}(t_K) | \hat{S}_z^{(s)}(\mathrm{t})\right) dS_z^{(s)}(\mathrm{t})$$

$$Step1: S_z^{(s)}(\mathrm{t}) \sim f\left(S_z^{(s)}(t_K) | \hat{S}_z^{(s)}(\mathrm{t})\right)$$

$$Step2: \hat{p}_z^{(s)} \sim f\left(\hat{p}_z^{(s)} | \hat{S}_z^{(s)}(\mathrm{t}), S_z^{(s)}(\mathrm{t})\right)$$

Bivariate binomial model for prognosis MA: MI

$$Step1: S_z^{(s)}(\mathrm{t}) \left| \hat{S}_z^{(s)}(\mathrm{t}) \sim f\left(S_z^{(s)}(\mathrm{t}) | \hat{S}_z^{(s)}(\mathrm{t})\right)\right.$$

$$\hat{S}_z^{(s)}(\mathrm{t}) \sim N\left(S_z^{(s)}(\mathrm{t}), \frac{\sigma_z^2(\mathrm{t})}{n_z^{(s)}}\right) \qquad \sigma_z^2(\mathrm{t}): \text{Greenwood variance}$$

$$S_z^{(s)}(\mathrm{t}) | \hat{S}_z^{(s)}(\mathrm{t}) \sim N\left(\hat{S}_z^{(s)}(\mathrm{t}), \frac{\sigma_z^2(\mathrm{t})}{n_z^{(s)}}\right) \text{ with a vague prior}$$

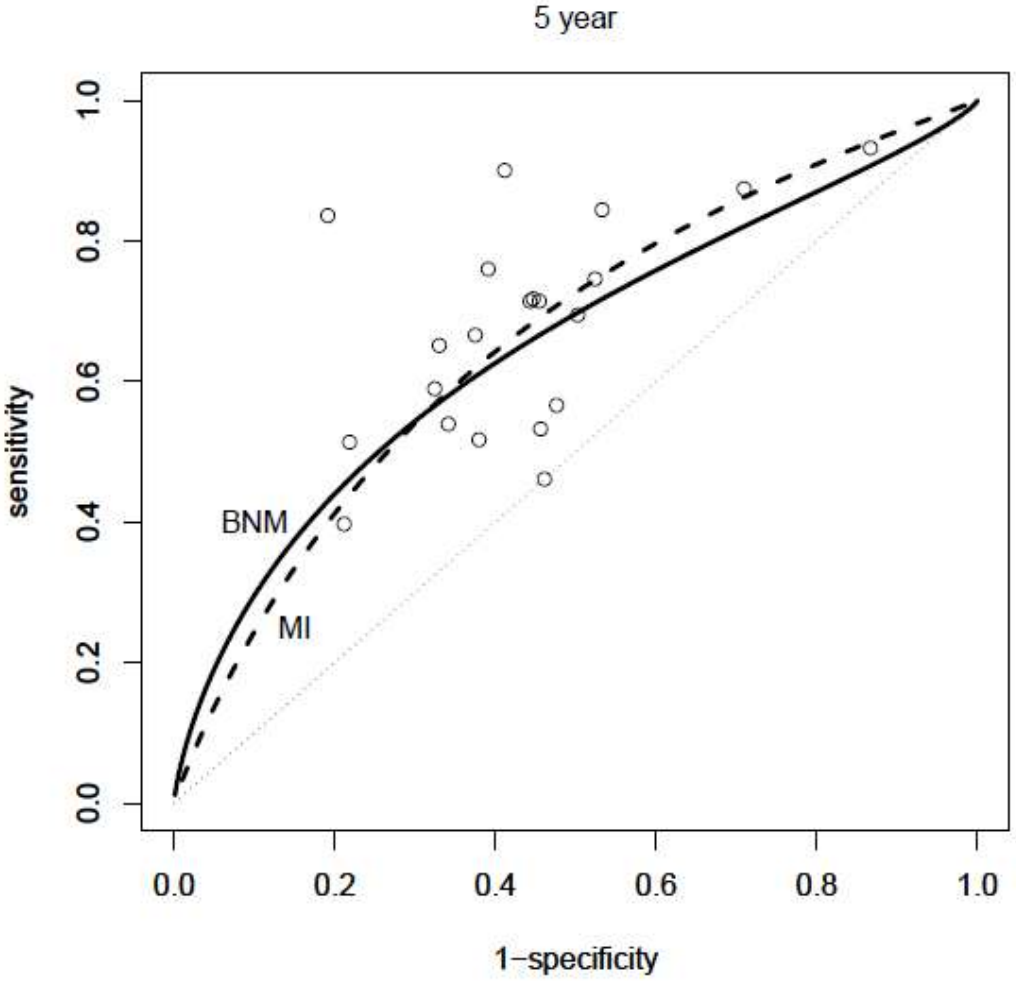
$$Step2: \hat{p}_z^{(s)} \left| \hat{S}_z^{(s)}(\mathrm{t}), S_z^{(s)}(\mathrm{t}) \sim f\left(\hat{p}_z^{(s)} | \hat{S}_z^{(s)}(\mathrm{t}), S_z^{(s)}(\mathrm{t})\right)\right.$$

$$\begin{pmatrix} \hat{p}_z^{(s)} \\ \hat{S}_z^{(s)}(\mathrm{t}) \end{pmatrix} \sim N\left(\begin{pmatrix} S_z^{(s)}(\mathrm{t}) \\ S_z^{(s)}(\mathrm{t}) \end{pmatrix}, \frac{1}{n_{z+}^{(s)}} \begin{pmatrix} u_z^{(s)}(\mathrm{t}), u_z^{(s)}(\mathrm{t}) \\ u_z^{(s)}(\mathrm{t}), \sigma_z^2(\mathrm{t}) \end{pmatrix}\right)$$

$$u_z^{(s)}(\mathrm{t}) = S_z^{(s)}(\mathrm{t})\left\{1 - S_z^{(s)}(\mathrm{t})\right\} \text{ :bionomial variance}$$

$$\hat{p}_z^{(s)} | \hat{S}_z^{(s)}(\mathrm{t}), S_z^{(s)}(\mathrm{t}) \sim normal$$

Application to Ki-67 data



Number of imputation :
$K = 10$

t	BMM	MI	Combesecure et al.(2015)
3	0.66 (0.61, 0.71)	0.65 (0.61, 0.69)	0.63 (0.57, 0.71)
5	0.65 (0.61, 0.69)	0.66 (0.62, 0.69)	0.64 (0.56, 0.74)

Assessing potential impacts of publication bias (PB)

Publication bias (PB) in diagnosis/prognosis MA

- Publication bias is a serious concern in validity of MA.
- For MA of intervention studies, various methods for PB
  - graphical methods
  - sensitivity analysis
- Simple graphical procedures are hard to apply to multivariate MA.
- Sensitivity analysis method is an objective framework to address PB.

Sensitivity analysis methods for MA of intervention studies

Various sensitivity analysis methods are available

Parametric approach:

- Heckman-type selection function (Copas 1999, JRSS-C; Copas and Shi 2000, Biostatistics)
- t-statistic based selection function (Copas 2013, JRSS-C)

Nonparametric approach:

- Nonparametric worst-case bound (Copas and Jackson 2004, Biometrics)

Sensitivity analysis methods for diagnosis MA

Several extensions have been made.

Bivariate Gaussian model;

- Piao et al. (2019, SMMR): Heckman-type
- Li et al. (2021, Metrika): Heckman-type
- Zhou, Huang and Hattori (2023, SiM): t-statistic type
- Zhou, Huang and Hattori (in preparation): nonparametric worst-case bound

Bivariate binomial model;

- Hattori and Zhou (2018, SiM): Heckman-type

Sensitivity analysis methods for prognosis MA

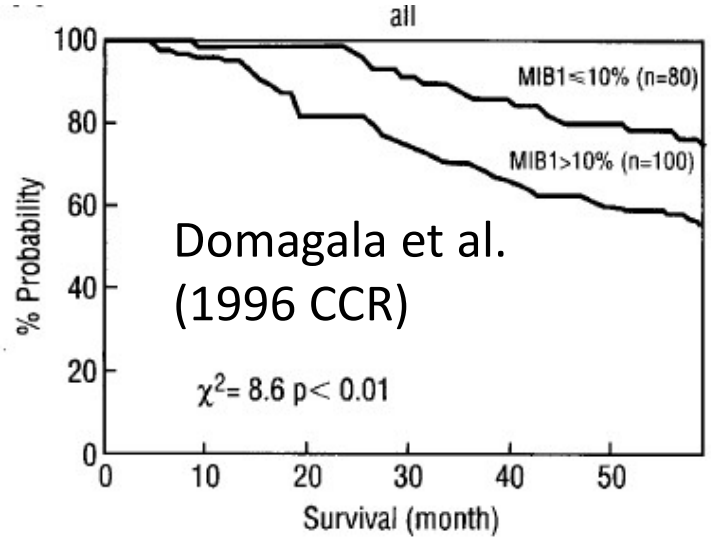
Zhou, Huang and Hattori (arXiv:2305.19741 [stat.ME])

- Results of the logrank test would be responsible for publication.

Selection function:

$\mathrm{P}(pubished|data) = \Phi(\alpha + \beta \times Z)$

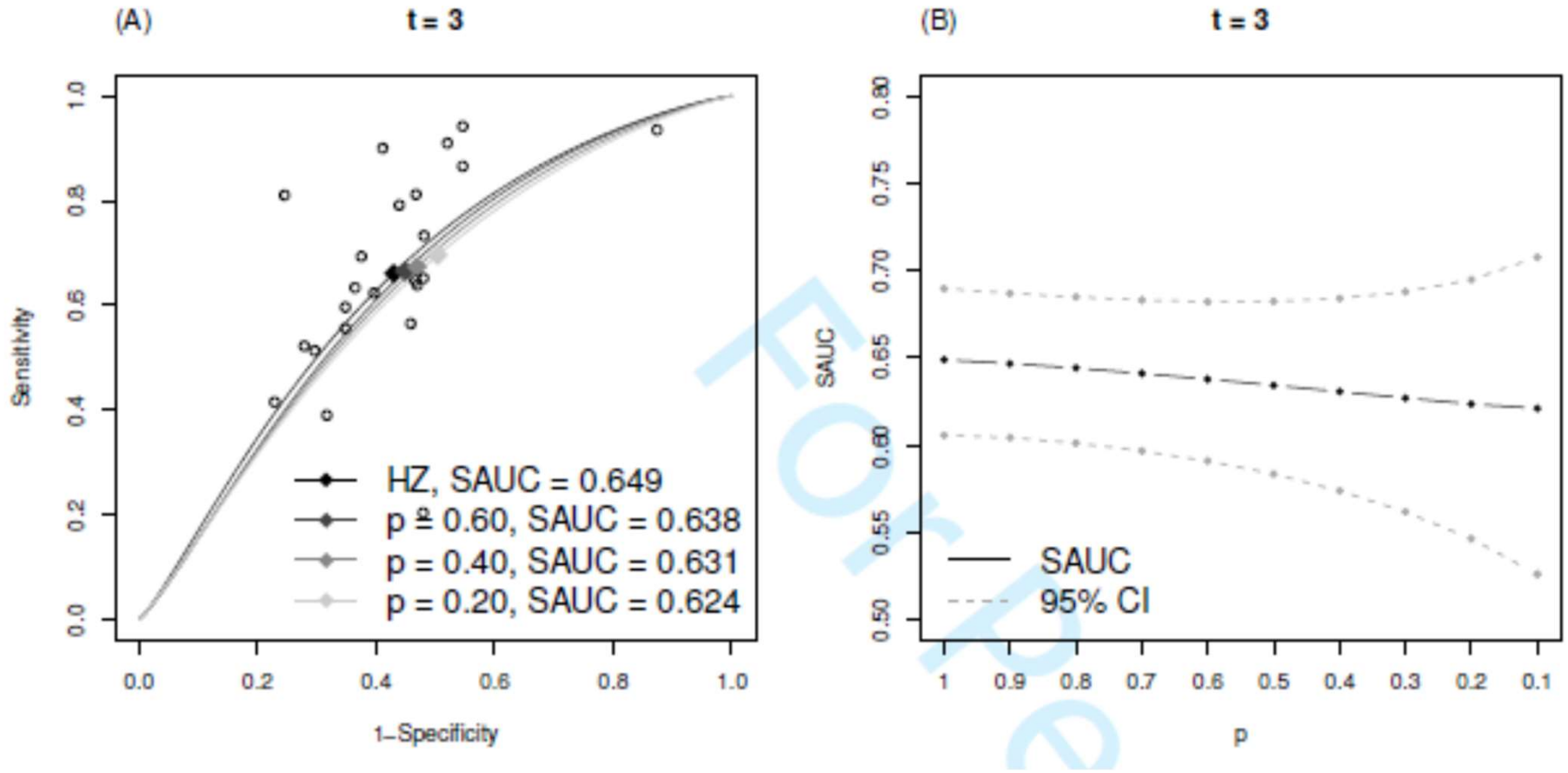
Z: the logrank test statistics



- Estimate time-dependent SROC under this selection process via (conditional) MLE
- Evaluate how the estimates may change as the marginal publication probability p=P$(pubished)$ decreases.

Sensitivity analysis methods for prognosis MA

Application to Ki-67 data



Concluding remarks

We reviewed recent methodological development of prognosis MA

- Meta-analytic version of time-dependent ROC
- Meta-analytic C-index for a survival outcome (Hattori and Zhou 2022, SiM)

- Software is under development and will be opened.

Thanks so much for your kind attention!