301154

A NEW METHOD TO ESTIMATE TREATMENT EFFECT AND CLASSIFIER FOR LATENT SUBGROUP

2020 ASA BIOPHARMACEUTICAL SECTION REGULATORY-INDUSTRY STATISTICS WORKSHOP

FRIDAY, SEPTEMBER 25 AT 11:45 A.M. – 12:45 P.M. EDT.

Kohei Uemura[1], Aya Kuchiba[1], Shogo Nomura[1]

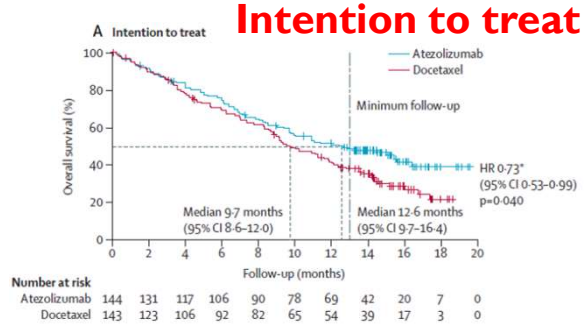
1: Department of Biostatistics and Bioinformatics, University of Tokyo
2: Biostatistics Division, CRAS, National Cancer Center

Introduction

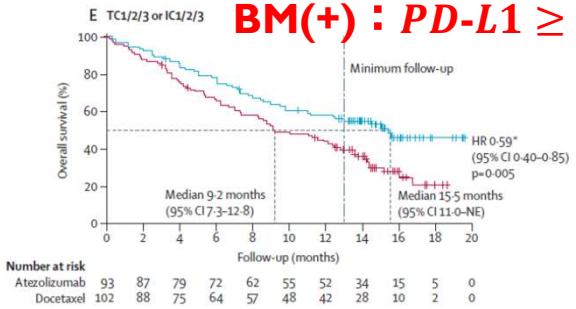
- Recently, immune checkpoint inhibitors of PD1/PD-L1 have been actively developed as cancer treatment.
- How to define target population: BM-positive ?
  - Those whose expression level of the biomarker (BM) are over than a cutoff value
- A problem is that
  - A currently used subgroup classifier based on PD1/PD-L1 expression level may be imperfect to select target patients
    - ✓ E.g. Atezolizumab may be effective in some patients regardless of PD-L1 negative according to a RCT: Atezolizumab vs. docetaxel for NSCLC (POPLAR)
- So there would exist a better classifier for treatment selection of PD1/PDL1 inhibitors
- Examinations of how far PD1/PD-L1 classifier departs from an ideal one require information on latent subgroup membership

Lancet 2016; 387: 1837–46

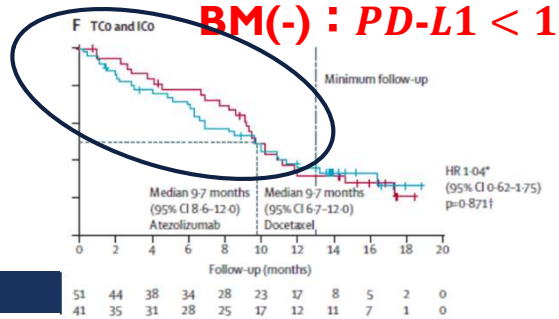
Intention to treat



BM(+) : $PD\text{-}L1 \geq 1\%$

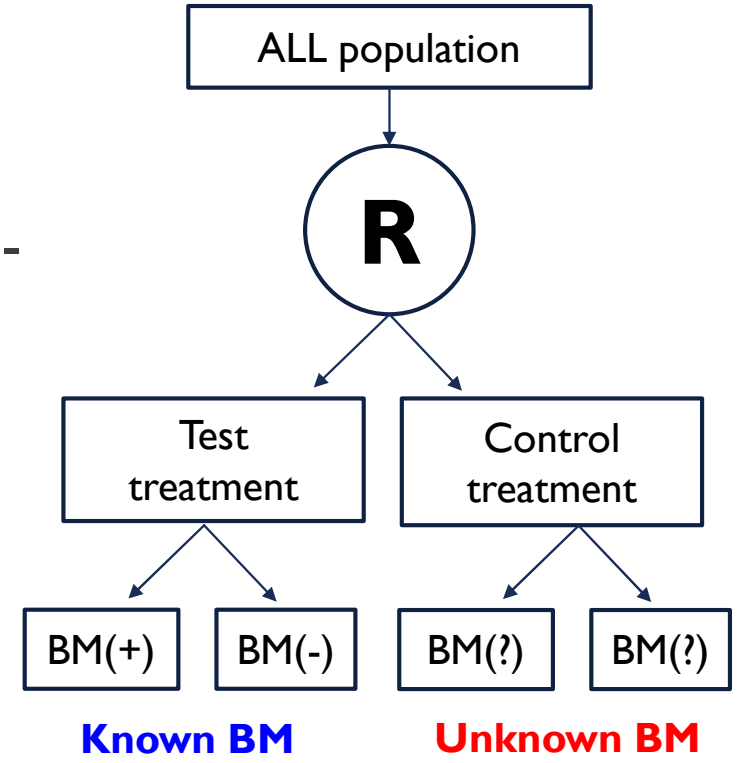


BM(-) : $PD\text{-}L1 < 1\%$



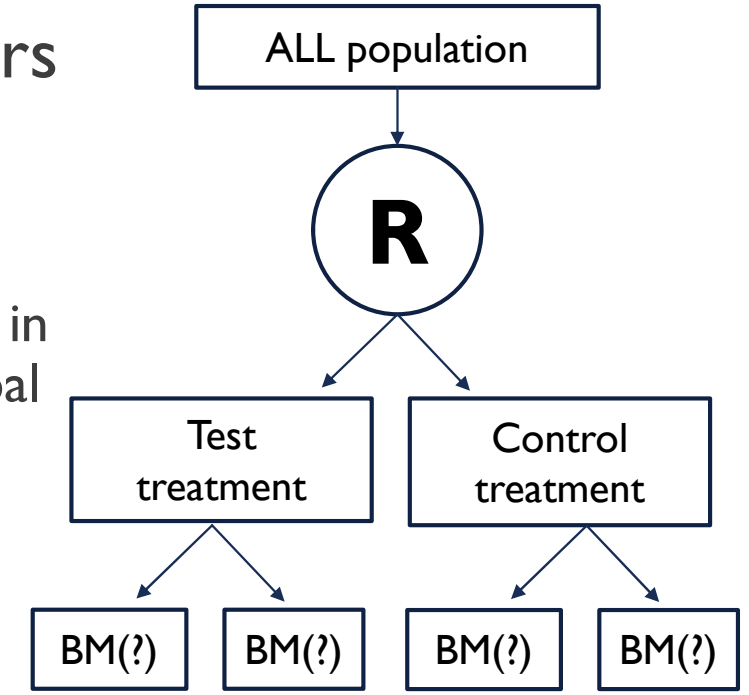
An existing method to estimate latent subgroup

- Altstein and Li, Biometrics. 2013; 69:52-61
  - Data from a randomized clinical trial
  - Utilize framework of principal stratification in non-compliance setting with partially known principal strata under monotonicity assumption
  - Estimate latent subgroups where latent subgroup membership is known for those who are assigned to one of the two arms in clinical trial.



A challenge to estimation of latent subgroup

- In situation of immune checkpoint inhibitors
  - latent subgroup membership is now unknown for both treatment arms
    - ✓ Can not utilize framework of principal stratification in non-compliance setting with partially known principal strata under monotonicity assumption
  - There are imperfect information (PD-L1 classifier) on latent subgroup membership



Unknown latent subgroup member -ships in both arms

Objective

- Propose a new statistical method
  - To estimate latent subgroup indicators under the situation where latent subgroup memberships are unknown for both treatment arms
  - Our methods utilize observed PD1/PD-L1 classifier as prior information and update individual status based on Bayes theorem according to survival outcome data
- Simulation study
  - Show preliminary results of estimation by proposed method with large samples
  - Also calculate C-index to assess how estimated subgroup indicators predict latent subgroup memberships

Notation

- T_i : Survival time of patient i
- C_i : Right censoring time of patient i
- $X_i = \min(T_i, C_i)$: Observed time of patient i
- $\delta_i = I(T_i \leq C_i)$: Indicator variable of event of patient i
- $R_i = 1$: Experimental treatment, $R_i = 0$: Control treatment
- G_i : Latent subgroup membership of patient i
  - $G_i = 1$: Target population of PD1/PD-L1 inhibitors, $G_i = 0$: Non-target population
- $\boldsymbol{Z}_i = \left(Z_{i1}, \ldots, Z_{iq}\right)'$: Observed covariate of patient i

AFT(Accelerate Failure Time) mixture model

- We use AFT mixture model
  - To express different treatment effects across unknwon true classifier G_i

$$Pr(G_i = 1) = p$$

$$\log(T_i|R_i, G_i, \boldsymbol{Z}_i) = \alpha_1 R_i + \alpha_2 G_i + \alpha_3 R_i G_i + \boldsymbol{Z}_i'\boldsymbol{\beta} + \sigma\varepsilon_i$$

  - p : parameter of proportion of a latent subgroup of $G_i = 1$
  - α_1: parameter of treatment effect for non-target population of $G_i = 0$
  - α_2: parameter of prognostic effect of target population of $G_i = 1$
  - α_3: parameter of predictive treatment effect for target population of $G_i = 1$
  - $\boldsymbol{\beta}$: parameter vector of intercept and effects of covariates
  - σ : scale parameter, ε_i : random error following extreme value distribution G(0, 1)

Estimation by EM algorithm

- Construction of likelihood based on the AFT mixture model
  - Each patient contributes to likelihood as both members of target and non-target population using estimated classifier $\hat{G}_i^{(m)}$

$$L\left(\alpha_1, \alpha_2, \alpha_3, \boldsymbol{\beta} \middle| R_i, \hat{G}_i^{(m)}, \boldsymbol{Z}_i\right)$$

$$= \prod_{i=1}^{n} \left[pf(X_i|R_i, G_i = 1, \boldsymbol{Z}_i)^{\delta_i} S(X_i|R_i, G_i = 1, \boldsymbol{Z}_i)^{1-\delta_i}\right]^{\hat{G}_i^{(m)}} \cdot \left[(1-p)f(X_i|R_i, G_i = 0, \boldsymbol{Z}_i)^{\delta_i} S(X_i|R_i, G_i = 0, \boldsymbol{Z}_i)^{1-\delta_i}\right]^{1-\hat{G}_i^{(m)}}$$

How to estimate $\hat{G}_i^{(m)}$?

- Bayes theorem
  - If T_i (survival time of patient i) is observed: $\delta_i = 1$

$$\hat{G}_i^{(m)} = \frac{Pr\left(T_i = x_i \middle| R_i, G_i = 1, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right) Pr\left(G_i = 1 \middle| \boldsymbol{\Theta}^{(m)}\right)}{\sum_{g \in \{0,\,1\}} Pr\left(T_i = x_i \middle| R_i, G_i = g, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right) Pr\left(G_i = g \middle| \boldsymbol{\Theta}^{(m)}\right)}$$

$$= \frac{p^{(m)} f\left(x_i \middle| R_i, G_i = 1, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right)}{p^{(m)} f\left(x_i \middle| R_i, G_i = 1, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right) + (1 - p^{(m)}) f\left(x_i \middle| R_i, G_i = 0, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right)}$$

Note that $\boldsymbol{\Theta}^{(m)} = \alpha_1^{(m)}, \alpha_2^{(m)}, \alpha_3^{(m)}, \boldsymbol{\beta}^{(m)}$

How to estimate $\hat{G}_i^{(m)}$?

- Bayes theorem
  - If T_i (survival time of patient i) is censored: $\delta_i = 0$

$$\hat{G}_i^{(m)} = \frac{Pr\left(T_i > x_i \middle| R_i, G_i = 1, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right) Pr\left(G_i = 1 \middle| \boldsymbol{\Theta}^{(m)}\right)}{\sum_{g \in \{0,\,1\}} Pr\left(T_i > x_i \middle| R_i, G_i = g, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right) Pr\left(G_i = g \middle| \boldsymbol{\Theta}^{(m)}\right)}$$

$$= \frac{p^{(m)} f\left(x_i \middle| R_i, G_i = 1, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right)}{p^{(m)} S\left(x_i \middle| R_i, G_i = 1, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right) + (1 - p^{(m)}) S\left(x_i \middle| R_i, G_i = 0, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right)}$$

Note that $\boldsymbol{\Theta}^{(m)} = \alpha_1^{(m)}, \alpha_2^{(m)}, \alpha_3^{(m)}, \boldsymbol{\beta}^{(m)}$

Proposed method

- Using information of observed classifier $PD\text{-}L1_i \in \{0, 1\}$
  - E-step： In case of $\delta_i = 1$

$$\hat{G}_i^{(m)} = \frac{\Pr[G_i{=}1|PD\text{-}L1_i]^{(m)} f\left(x_i \middle| R_i, G_i = 1, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right)}{\Pr[G_i{=}1|PD\text{-}L1_i]^{(m)} f\left(x_i \middle| R_i, G_i = 1, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right) + \left(1 - \Pr[G_i{=}1|PD\text{-}L1_i]^{(m)}\right) f\left(x_i \middle| R_i, G_i = 0, \boldsymbol{Z}_i, \boldsymbol{\Theta}^{(m)}\right)}$$

$$\Pr[G_i = 1|PD\text{-}L1_i = 1]^{(m)} = \frac{p^{(m)} se^{(m)}}{p^{(m)} se^{(m)} + (1 - p^{(m)})(1 - sp^{(m)})}$$

$$\Pr[G_i = 1|PD\text{-}L1_i = 0]^{(m)} = \frac{p^{(m)}(1 - se^{(m)})}{p^{(m)}(1 - se^{(m)}) + (1 - p^{(m)}) sp^{(m)}}$$

Note that $se = \Pr[PD\text{-}L1_i = 1|G_i = 1], sp = \Pr[PD\text{-}L1_i = 0|G_i = 0]$

Proposed method (cont')

- Using information of observed classifier $PD\text{-}L1_i \in \{0, 1\}$
  - M-step：

$$\alpha_1^{(m+1)}, \alpha_2^{(m+1)}, \alpha_3^{(m+1)}, \boldsymbol{\beta}^{(m+1)} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \left[L\left(p \middle| X_i, \hat{G}_i^{(m)}\right) L\left(\alpha_1, \alpha_2, \alpha_3, \boldsymbol{\beta} \middle| X_i, \hat{G}_i^{(m)}\right) \right]$$

$$p^{(m+1)} = \underset{p}{\operatorname{argmax}} \left[L\left(p \middle| X_i, \hat{G}_i^{(m)}\right) L\left(\alpha_1, \alpha_2, \alpha_3, \boldsymbol{\beta} \middle| X_i, \hat{G}_i^{(m)}\right) \right] = \frac{1}{n} \sum_{i=1}^{n} \hat{G}_i^{(m)}$$

$$se^{(m+1)} = \Pr[PD\text{-}L1_i = 1 | G_i = 1] = \frac{\frac{1}{n} \sum_{i=1}^{n} PD\text{-}L1_i \hat{G}_i^{(m)}}{p^{(m)}}$$

$$sp^{(m+1)} = \Pr[PD\text{-}L1_i = 0 | G_i = 0] = \frac{\frac{1}{n} \sum_{i=1}^{n} (1 - PD\text{-}L1_i)(1 - \hat{G}_i^{(m)})}{1 - p^{(m)}}$$

Simulation study

$$Pr(G_i = 1) = p$$
$$\log(T_i|R_i, G_i, \boldsymbol{Z}_i) = \alpha_1 R_i + \alpha_2 G_i + \alpha_3 R_i G_i + \sigma\varepsilon_i$$

- Considering 6 scenarios
  - $p = 0.6$: initial values are 0.2, 0.4, 0.6, 0.8
- In hazard ratio scale
  - $\exp(-\alpha_1)$: 0
  - $\exp(-\alpha_2)$: 5.0, 3.0, 1.5
  - $\exp(-\alpha_3)$: 0.1, 0.3, 0.5, 0.7
- n/group=10,000

Results of estimation without info. of observed classifier $PD\text{-}L1_i$

Large bias of estimates for small effect sizes of HR

	scenario[a)]	effect	Estimated HR	95%CI		Estimated p[b)]
①	0.2, 0.1, 5	treat	0.098	[0.094 ,	0.102]	0.588
		SubG	4.953	[4.783 ,	5.123]	.
②	0.2, 0.3, 5	treat	0.300	[0.289 ,	0.311]	0.593
		SubG	4.919	[4.750 ,	5.089]	.
③	0.2, 0.5, 5	treat	0.496	[0.478 ,	0.514]	0.586
		SubG	4.900	[4.730 ,	5.069]	.
④	0.2, 0.7, 5	treat	0.699	[0.674 ,	0.725]	0.576
		SubG	4.913	[4.744 ,	5.083]	.
⑤	0.2, 0.7, 3	treat	0.673	[0.647 ,	0.699]	0.520
		SubG	3.013	[2.909 ,	3.117]	.
	more restrict convergence criteria	treat	0.703	[0.678 ,	0.729]	0.597
		SubG	2.963	[2.860 ,	3.066]	.

a) Initial value for p, Treatment HR: $\exp(-\alpha_3)$,
Latent subgroup prognostic HR:$\exp(-\alpha_2)$.
b) True value of p (proportion of latent subgroup) is set at 0.6.

	scenario[a)]	effect	Estimated HR	95%CI		Estimated p[b)]
⑥	0.2, 0.7, 1.5	treat	0.389	[0.364 ,	0.413]	0.216
		SubG	1.771	[1.69 ,	1.852]	.
	0.4, 0.7, 1.5	treat	0.600	[0.572 ,	0.627]	0.402
		SubG	1.552	[1.494 ,	1.609]	.
	0.6, 0.7, 1.5	treat	0.710	[0.684 ,	0.737]	0.599
		SubG	1.628	[1.571 ,	1.685]	.
	0.8, 0.7, 1.5	treat	0.775	[0.75 ,	0.8]	0.794
		SubG	1.864	[1.787 ,	1.94]	.

a) Initial value for p, Treatment HR: $\exp(-\alpha_3)$,
Latent subgroup prognostic HR:$\exp(-\alpha_2)$.
b) True value of p (proportion of latent subgroup) is set at 0.6.

Results of estimation with info. of observed classifier $PD\text{-}L1_i$

No or little bias of estimates for small effect sizes of HR

	scenario[a)]	effect	Estimated HR	95%CI		Estimated p[b)]
①	0.2, 0.1, 5	treat	0.100	[0.095 ,	0.104]	0.593
		SubG	4.967	[4.795 ,	5.140]	.
②	0.2, 0.3, 5	treat	0.300	[0.289 ,	0.311]	0.589
		SubG	4.923	[4.751 ,	5.096]	.
③	0.2, 0.5, 5	treat	0.505	[0.487 ,	0.523]	0.596
		SubG	4.935	[4.763 ,	5.107]	.
④	0.2, 0.7, 5	treat	0.707	[0.682 ,	0.733]	0.601
		SubG	4.952	[4.779 ,	5.124]	.
⑤	0.2, 0.7, 3	treat	0.707	[0.681 ,	0.732]	0.597
		SubG	2.957	[2.854 ,	3.060]	.
⑥	0.2, 0.7, 1.5	treat	0.678	[0.650 ,	0.705]	0.535
		SubG	1.496	[1.444 ,	1.549]	.

a) Initial value for p, Treatment HR: $\exp(-\alpha_3)$,

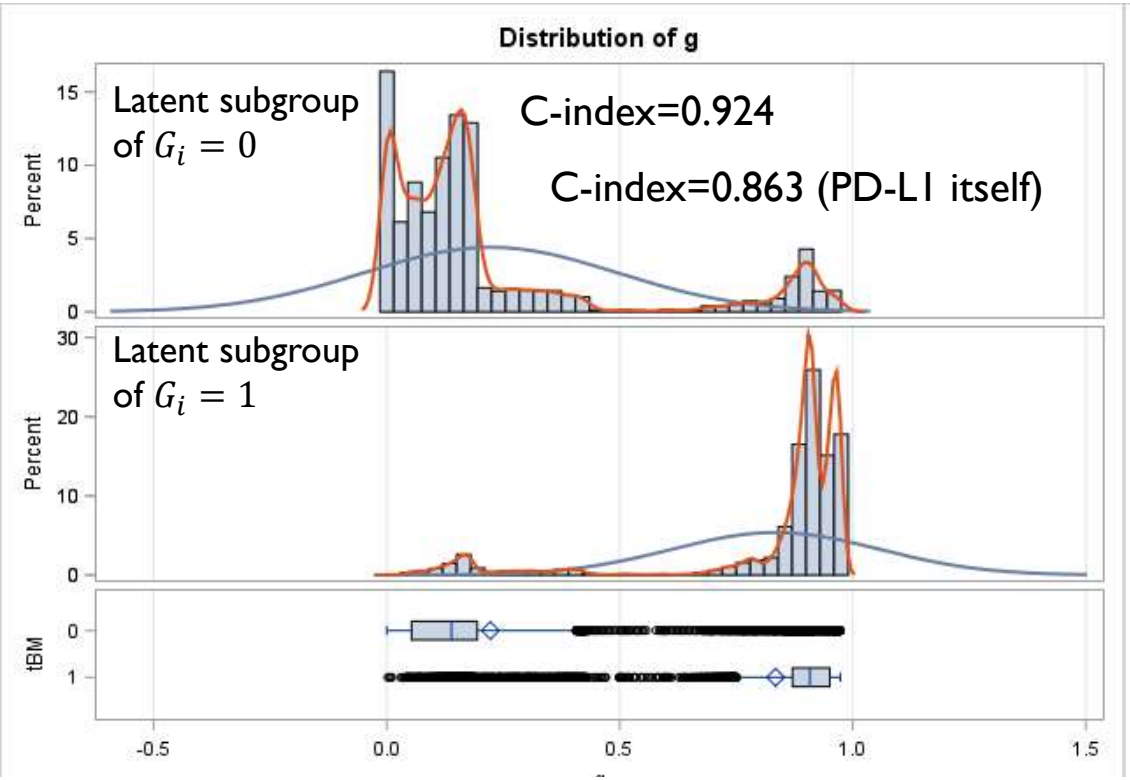
Latent subgroup prognostic HR:$\exp(-\alpha_2)$.

b) True value of p (proportion of latent subgroup) is set at 0.6.

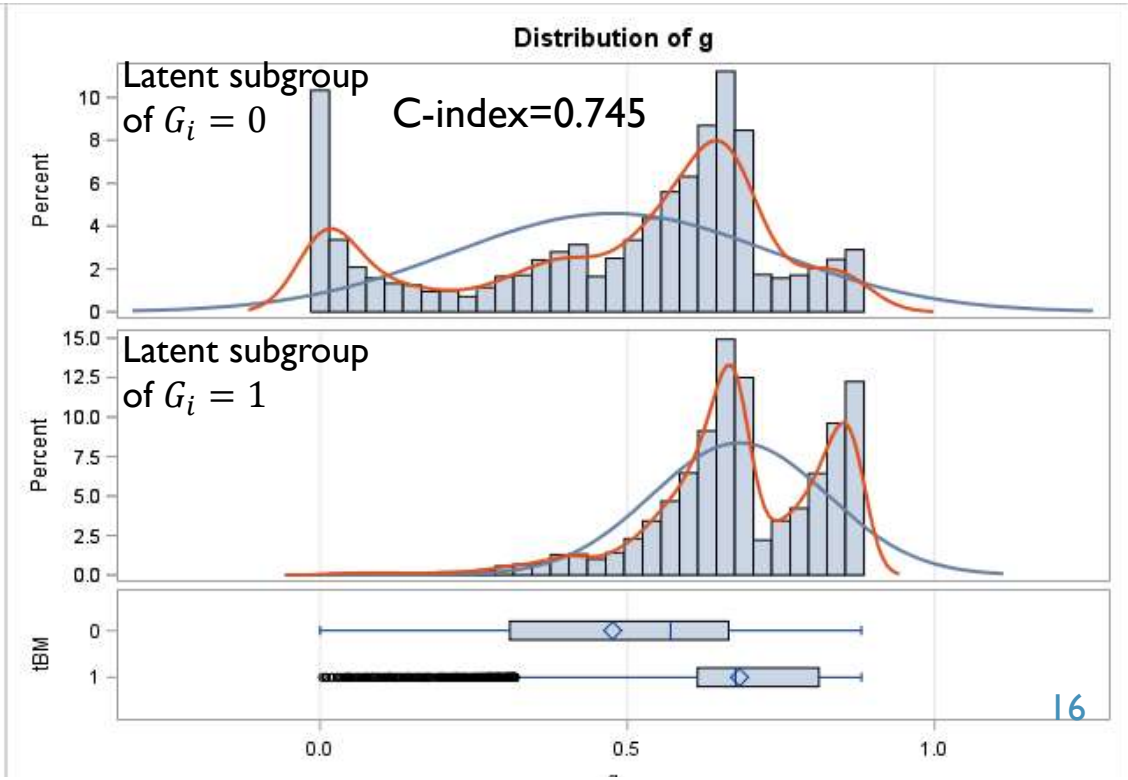
Distribution of $\hat{G}_i$: scenario ②

C-index is for prediction of true G_i

Estimation using prior information of PD-L1 (proposed)



Estimation not using prior



Summary

- We propose a new statistical method
  - To estimate latent subgroup indicators under the situation where latent subgroup memberships are unknown for both treatment arms
  - Our methods utilize observed PD1/PD-L1 classifier as prior information
- Our method can estimate treatment effects and memberships for latent subgroups
  - Without major bias even in case of small effect sizes concerning treatment and subgroup
  - With higher values of C-index for prediction of each patient's latent memberships
- Our method will be useful for examinations of how far PD1/PD-L1 classifier departs from an ideal one
  - PD1/PD-L1 expression level is relating to activity of effector phase of T-cell, but other immune system activity might mainly or additionally relate to cancer progression including priming phase
  - Uncertainty of cutoff values of BMs

Please contact us if any. Email : Kohei.Uemura@iii.u-tokyo.ac.jp