Biometrics 000, 000–000 DOI: 000
000 0000

Supporting Information for

"Causal Meta-Analysis by Integrating Multiple Observational Studies with Multivariate Outcomes"

by Subharup Guha and Yi Li

This paper has been submitted for consideration for publication in *Biometrics*

A. Proofs of theoretical results

A.1 *Theorem 1*

Applying equation (3),

$$\begin{aligned}
\mathbb{E}_+\big[\rho^2(S,Z,\mathbf{X})\big] &= \frac{1}{\big(\mathbb{E}_+[\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{X})]\big)^2}\mathbb{E}_+\left[\frac{\gamma_S^2\,\theta_Z^2\,\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}^2(\mathbf{X})}{\delta_{SZ}^2(\mathbf{X})}\right] \\
&= \frac{1}{\big(\mathbb{E}_+[\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{X})]\big)^2}\mathbb{E}_+\left[\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}^2(\mathbf{X})\sum_{s=1}^{J}\sum_{z=1}^{K}\frac{\gamma_s^2\theta_z^2}{\delta_{sz}^2(\mathbf{X})}\right] \\
&= \frac{\mathbb{E}_+\left[\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}^2(\mathbf{X})/\breve{\eta}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{X})\right]}{\big(\mathbb{E}_+[\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{X})]\big)^2} \\
&\geqslant \frac{1}{\mathbb{E}_+\big[\breve{\eta}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{X})\big]} \qquad \text{by Sedrakyan's inequality.}
\end{aligned}$$

The lower bound is attained when the tilting function $\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x})$ satisfies $\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x})/\sqrt{\breve{\eta}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x})} \propto \sqrt{\breve{\eta}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x})}$, i.e., when $\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x}) \propto \breve{\eta}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x})$. Evaluating equation (3) with $\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x}) \propto \breve{\eta}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x})$, we obtain the optimal fixed-$(\boldsymbol{\gamma},\boldsymbol{\theta})$ pseudo-population's weights. These weights are uniformly bounded for $(s,z) \in \Sigma$ and $\mathbf{x} \in \mathcal{X}$:

$$\begin{aligned}
\tilde{\rho}(s,z,\mathbf{x}) &= \frac{\gamma_s\theta_z\,\breve{\eta}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{x})}{\delta_{sz}(\mathbf{x})} \\
&= \frac{1}{\gamma_s\theta_z}\left(\frac{\gamma_s^2\theta_z^2/\delta_{sz}(\mathbf{x})}{\sum_{t=1}^{J}\sum_{u=1}^{K}\gamma_t^2\theta_u^2/\delta_{tu}(\mathbf{x})}\right) \\
&\leqslant \gamma_s^{-1}\theta_z^{-1} \leqslant \big(\min_{(s,z)\in\Sigma}\gamma_s\theta_z\big)^{-1},
\end{aligned}$$

which is finite.

A.2 *Theorem 2 details and proof*

(1) **Consistency**:

Proof. Consider the mth component of random vector $\bar{\mathbf{\Phi}}_z$:

$$\begin{aligned}
\bar{\Phi}_{zm} &= \frac{\sum_{i=1}^{N} \tilde{\rho}_i \, \Phi_m(\mathbf{Y}_i) \, \mathcal{I}(Z_i = z)}{\sum_{i=1}^{N} \tilde{\rho}_i \, \mathcal{I}(Z_i = z)} \\
&\xrightarrow{p} \frac{\mathbb{E}_+\big[\tilde{\rho}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(S, Z, \mathbf{X}) \, \Phi_m(\mathbf{Y}) \, \mathcal{I}(Z = z)\big]}{\mathbb{E}_+\big[\tilde{\rho}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(S, Z, \mathbf{X}) \, \mathcal{I}(Z = z)\big]} \\
&= \frac{\mathbb{E}\big[\Phi_m(\mathbf{Y}) \, \mathcal{I}(Z = z)\big]}{\mathbb{E}\big[\mathcal{I}(Z = z)\big]} \\
&= \mathbb{E}\big[\Phi_m(\mathbf{Y}) \,\big|\, Z = z\big] \\
&= \mathbb{E}\big[\Phi_m(\mathbf{Y}^{(z)})\big]
\end{aligned}$$

for the covariate-balanced pseudo-population.

(2) **Asymptotic normality**:

(a) **Known MPS**

Proof. Estimator (7) is a solution to the following estimating equation with respect to $\boldsymbol{\lambda}^{(z)}$:

$$\sum_{i=1}^{N} \mathbb{S}_{iz}^{(\lambda)} = \mathbf{0}_M,$$

where $\mathbb{S}_{iz}^{(\lambda)} := \tilde{\rho}_i \, \mathcal{I}(Z_i = z) \big(\mathbf{\Phi}(\mathbf{Y}_i) - \boldsymbol{\lambda}^{(z)}\big)$ is a vector of length M. Define $\mathbf{A}_3^{(z)} := \mathbb{E}_+\big[\partial \mathbb{S}_{iz}^{(\lambda)} / \partial \boldsymbol{\lambda}^{(z)\prime}\big]$ which equals $-\mathbb{E}_+\big(\tilde{\rho}_{\boldsymbol{\gamma},\boldsymbol{\theta}}(S, Z, \mathbf{X}) \; \mathcal{I}(Z = z)\big)\mathbf{I}_M = -\theta_z \mathbb{E}_+[\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{X})]\mathbf{I}_M$. Define $\mathbf{B}_3^{(z)} := \mathbb{E}_+\big[\mathbb{S}_{iz}^{(\lambda)} \mathbb{S}_{iz}^{(\lambda)\prime}\big]$ which equals $\mathbb{E}_+^2[\eta_{\boldsymbol{\gamma},\boldsymbol{\theta}}(\mathbf{X})] \, \mathbf{\Sigma}_1^{(z)}$. Applying standard large-sample results for estimating equations, $\bar{\mathbf{\Phi}}_z$ is asymptotically normal and centered at $\boldsymbol{\lambda}^{(z)}$, with

$$\lim_{N \to \infty} N \mathrm{var}(\bar{\mathbf{\Phi}}_z) = \big(\mathbf{A}_3^{(z)}\big)^{-1} \mathbf{B}_3^{(z)} \big(\mathbf{A}_3^{(z)}\big)^{-1} = \mathbf{\Sigma}_1^{(z)},$$

defined in the theorem statement.

(b) **Estimated MPS**

Proof. Let the MPSs be concatenated as $\boldsymbol{\delta}(\mathbf{x}) = \text{vec}\big\{\hat{\delta}_{s^*z^*}(\mathbf{x}) : (s^*, z^*) \neq (1,1)\big\}$. To make explicit their dependence on $\boldsymbol{\omega}$, we write the matrices $\mathbf{A}_3^{(z)}$ and $\mathbf{B}_3^{(z)}$, defined in the previous part, as $\mathbf{A}_3^{(z)}(\boldsymbol{\omega})$ and $\mathbf{B}_3^{(z)}(\boldsymbol{\omega})$ respectively. Extending the analytical approaches of Mao et al. (2019) and Zeng et al. (2023), we observe that estimator (7) is a solution to the following estimating equation with respect to $\boldsymbol{\lambda}^{(z)}$:

$$\sum_{i=1}^{N} \begin{bmatrix} \mathbb{S}_{iz}^{(\omega)} \\ \mathbb{S}_{iz}^{(\lambda)} \end{bmatrix} = \mathbf{0}_{(JK-1)p+M},$$

where $\mathbb{S}_{iz}^{(\omega)} := \text{vec}\Big\{\mathbf{x}_i\Big(\mathcal{I}(s_i = s^*, z_i = z^*) - \delta_{s^*z^*}(\mathbf{x})\Big) : (s^*, z^*) \neq (1,1)\Big\}$ is a vector of length $(JK-1)p$. Let $\mathbf{A}_1(\boldsymbol{\omega}) := \mathbb{E}_+\big[\partial \mathbb{S}_{iz}^{(\omega)}/\partial \boldsymbol{\omega}'\big]$ and $\mathbf{A}_2^{(z)}(\boldsymbol{\omega}) := \mathbb{E}_+\big[\partial \mathbb{S}_{iz}^{(\lambda)}/\partial \boldsymbol{\omega}'\big]$. Also, let $\mathbf{B}_1(\boldsymbol{\omega}) := \mathbb{E}_+\big[\mathbb{S}_{iz}^{(\omega)}\mathbb{S}_{iz}^{(\omega)\prime}\big]$ and $\mathbf{B}_2^{(z)}(\boldsymbol{\omega}) := \mathbb{E}_+\big[\mathbb{S}_{iz}^{(\omega)}\mathbb{S}_{iz}^{(\lambda)\prime}\big]$. Now, define the matrix

$$\mathbf{A}^{(z)}(\boldsymbol{\omega}) := \begin{bmatrix} \mathbb{E}_+\left(\frac{\partial \mathbb{S}_{iz}^{(\omega)}}{\partial \boldsymbol{\omega}'}\right) & \mathbb{E}_+\left(\frac{\partial \mathbb{S}_{iz}^{(\omega)}}{\partial \boldsymbol{\lambda}^{(z)\prime}}\right) \\ \mathbb{E}_+\left(\frac{\partial \mathbb{S}_{iz}^{(\lambda)}}{\partial \boldsymbol{\omega}'}\right) & \mathbb{E}_+\left(\frac{\partial \mathbb{S}_{iz}^{(\lambda)}}{\partial \boldsymbol{\lambda}^{(z)\prime}}\right) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1(\boldsymbol{\omega}) & \mathbf{0}_{(JK-1)p\times M} \\ \mathbf{A}_2^{(z)}(\boldsymbol{\omega}) & \mathbf{A}_3^{(z)}(\boldsymbol{\omega}) \end{bmatrix}.$$

Hence, writing $\mathbf{G}^{(z)}(\boldsymbol{\omega})\big) = \big(\mathbf{A}^{(z)}(\boldsymbol{\omega})\big)^{-1}$, we have

$$\begin{aligned} \mathbf{G}^{(z)}(\boldsymbol{\omega}) &= \begin{bmatrix} \mathbf{A}_1^{-1}(\boldsymbol{\omega}) & \mathbf{0}_{(JK-1)p\times M} \\ -\big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1}\mathbf{A}_2^{(z)}(\boldsymbol{\omega})\big(\mathbf{A}_1(\boldsymbol{\omega})\big)^{-1} & \big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_1^{-1}(\boldsymbol{\omega}) & \mathbf{0}_{(JK-1)p\times M} \\ \mathbf{C}^{(z)}(\boldsymbol{\omega}) & \big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1} \end{bmatrix}, \end{aligned}$$

where $\mathbf{C}^{(z)}(\boldsymbol{\omega}) = -\big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1}\mathbf{A}_2^{(z)}(\boldsymbol{\omega})\big(\mathbf{A}_1(\boldsymbol{\omega})\big)^{-1}$ is a matrix of dimension $M\times(JK-1)p$.

And

$$\mathbf{B}^{(z)}(\boldsymbol{\omega}) := \begin{bmatrix} \mathbb{E}_+\big(\mathbb{S}_{iz}^{(\omega)}\mathbb{S}_{iz}^{(\omega)\prime}\big) & \mathbb{E}_+\big(\mathbb{S}_{iz}^{(\omega)}\mathbb{S}_{iz}^{(\lambda)\prime}\big) \\ \mathbb{E}_+\big(\mathbb{S}_{iz}^{(\lambda)}\mathbb{S}_{iz}^{(\omega)\prime}\big) & \mathbb{E}_+\big(\mathbb{S}_{iz}^{(\lambda)}\mathbb{S}_{iz}^{(\lambda)\prime}\big) \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{B}_1(\boldsymbol{\omega}) & \mathbf{B}_2^{(z)}(\boldsymbol{\omega}) \\ \big(\mathbf{B}_2^{(z)}(\boldsymbol{\omega})\big)' & \mathbf{B}_3^{(z)}(\boldsymbol{\omega}) \end{bmatrix}.$$

Again applying standard results for estimating equations, vector $(\hat{\boldsymbol{\omega}}, \bar{\boldsymbol{\Phi}}_z)$ is asymptotically normal, centered at $(\boldsymbol{\omega}, \boldsymbol{\lambda}^{(z)})$, and

$$\lim_{N\to\infty} N \operatorname{var}(\hat{\boldsymbol{\omega}}, \bar{\boldsymbol{\Phi}}_z)) = \mathbf{G}^{(z)}(\boldsymbol{\omega})\mathbf{B}^{(z)}(\boldsymbol{\omega})\big(\mathbf{G}(\boldsymbol{\omega})\big)',$$

whose lower-right block matrix gives

$$\begin{aligned} \lim_{N\to\infty} N \operatorname{var}(\bar{\boldsymbol{\Phi}}_z) &= \big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1}\mathbf{B}_3^{(z)}(\boldsymbol{\omega})\big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1} + \mathbf{C}^{(z)}(\boldsymbol{\omega})\mathbf{B}_1(\boldsymbol{\omega})(\mathbf{C}^{(z)}\big(\boldsymbol{\omega})\big)' \\ &\quad + \big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1}\big(\mathbf{B}_2^{(z)}(\boldsymbol{\omega})\big)'\big(\mathbf{C}^{(z)}(\boldsymbol{\omega})\big)' + \mathbf{C}^{(z)}(\boldsymbol{\omega})\mathbf{B}_2^{(z)}(\boldsymbol{\omega})\big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1} \\ &= \boldsymbol{\Sigma}_1^{(z)}(\boldsymbol{\omega}) + \mathbf{D}^{(z)}(\boldsymbol{\omega}) \\ &= \boldsymbol{\Sigma}_2^{(z)}(\boldsymbol{\omega}). \end{aligned}$$

In other words, the adjustment $\mathbf{D}^{(z)}(\boldsymbol{\omega})$ in the theorem statement has the expression

$$\mathbf{C}^{(z)}(\boldsymbol{\omega})\mathbf{B}_1(\boldsymbol{\omega})(\mathbf{C}^{(z)}\big(\boldsymbol{\omega})\big)' + \big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1}\big(\mathbf{B}_2^{(z)}(\boldsymbol{\omega})\big)'\big(\mathbf{C}^{(z)}(\boldsymbol{\omega})\big)' + \mathbf{C}^{(z)}(\boldsymbol{\omega})\mathbf{B}_2^{(z)}(\boldsymbol{\omega})\big(\mathbf{A}_3^{(z)}(\boldsymbol{\omega})\big)^{-1}.$$

A.3 *Corollary 1*

The result follows by applying the delta method.

B. Additional simulation and data analysis results

[Table 1 about here.]

Table 1 compares the asymptotic and empirical variances of weighted estimators of ATE for artificial datasets. Table 2 summarizes a subset of the demographic, clinicopathological, and biomarker measurements. For the FLEXOR, IC, and IGO pseudo-populations, Tables 4–6 respectively present the 95% confidence intervals of mRNA expression level correlation between the gene pairs for the two breast cancer subtypes.

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

References

Mao, H., Li, L., and Greene, T. (2019). Propensity score weighting analysis and treatment effect discovery. *Statistical methods in medical research* **28,** 2439–2454.

Zeng, S., Li, F., Hu, L., and Li, F. (2023). Propensity score weighting analysis of survival outcomes using pseudo-observations. *Statistica Sinica* **33,** 2161–2184.

Average SD	**Similarity scenarios**					
	Low			**High**		
	$\tilde{N} = 125$ subjects					
	FLEXOR	**IGO**	**IC**	**FLEXOR**	**IGO**	**IC**
Asymptotic	0.80	1.85	1.92	0.76	0.78	0.81
Bootstrap-based	0.91	1.32	1.40	0.73	0.75	0.75
	$\tilde{N} = 250$ subjects					
	FLEXOR	**IGO**	**IC**	**FLEXOR**	**IGO**	**IC**
Asymptotic	0.47	1.05	1.10	0.45	0.46	0.48
Bootstrap-based	0.57	0.89	0.94	0.47	0.48	0.48
	$\tilde{N} = 500$ subjects					
	FLEXOR	**IGO**	**IC**	**FLEXOR**	**IGO**	**IC**
Asymptotic	0.21	0.48	0.50	0.20	0.21	0.22
Bootstrap-based	0.26	0.46	0.49	0.20	0.21	0.21

Table 1: Comparison of the limiting theoretical and bootstrap-based standard deviations for estimating the mean group difference of the two groups in the simulation study. The displayed values are the averages over the 500 artificial datasets. See Section 4 of the paper for further explanation.

Table 2: Summary of some demographic, clinicopathological, and biomarker variables of the TCGA breast cancer studies. Shown in parentheses are percentages. **IGC**: International Genomics Consortium; **MSKCC**: Memorial Sloan Kettering Cancer Center; **Pittsburgh**: University of Pittsburgh; **Miami**: University of Miami.

	Walter Reed	IGC	MSKCC	Mayo Clinic	Pittsburgh	Roswell Park	Miami
TOTAL	92	38	35	57	117	81	30
Mean age at diagnosis	56.6	66.4	52.5	54.6	56.7	57.1	57.8
Race							
Asian	2 (2.2)	1 (2.6)	2 (5.7)	1 (1.8)	1 (0.9)	0 (0.0)	1 (3.3)
Black	23 (25.0)	12 (31.6)	5 (14.3)	1 (1.8)	9 (7.7)	14 (17.3)	7 (23.3)
White	67 (72.8)	25 (65.8)	28 (80.0)	55 (96.5)	107 (91.5)	67 (82.7)	22 (73.3)
Cancer in nearby lymph nodes	45 (48.9)	26 (68.4)	18 (51.4)	30 (52.6)	65 (55.6)	42 (51.9)	18 (60.0)
Mean percentage genome altered	28.9	19.4	32.5	30.6	25.8	28.2	29.9
Median year of diagnosis	2008	2011	2006	2006	2008	2009	2010
Menopause status type							
Type 1 or 2	2 (2.2)	3 (7.9)	1 (2.9)	2 (3.5)	9 (7.7)	2 (2.5)	0 (0.0)
Type 3	62 (67.4)	23 (60.5)	19 (54.3)	33 (57.9)	58 (49.6)	54 (66.7)	23 (76.7)
Type 4	28 (30.4)	0 (0.0)	15 (42.9)	22 (38.6)	25 (21.4)	25 (30.9)	4 (13.3)
Type 5	0 (0.0)	12 (31.6)	0 (0.0)	0 (0.0)	25 (21.4)	0 (0.0)	3 (10.0)
Cancer stage							
Stage I	18 (19.6)	2 (5.3)	4 (11.4)	11 (19.3)	37 (31.6)	21 (25.9)	4 (13.3)
Stage II	50 (54.3)	26 (68.4)	24 (68.6)	30 (52.6)	57 (48.7)	45 (55.6)	19 (63.3)
Stage III	20 (21.7)	10 (26.3)	7 (20.0)	16 (28.1)	22 (18.8)	15 (18.5)	7 (23.3)
Stage IV	4 (4.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.9)	0 (0.0)	0 (0.0)
Positive ER status	68 (73.9)	35 (92.1)	25 (71.4)	48 (84.2)	99 (84.6)	60 (74.1)	21 (70.0)
Positive PR status	55 (59.8)	27 (71.1)	24 (68.6)	37 (64.9)	88 (75.2)	55 (67.9)	18 (60.0)
Cancer type							
Infiltrating Ductal Carcinoma (IDC)	72 (78.3)	11 (28.9)	29 (82.9)	40 (70.2)	105 (89.7)	67 (82.7)	23 (76.7)
Infiltrating Lobular Carcinoma (ILC)	20 (21.7)	27 (71.1)	6 (17.1)	17 (29.8)	12 (10.3)	14 (17.3)	7 (23.3)
Mean mRNA expression of gene							
COL9A3	-0.07	0.22	0.00	-0.12	-0.14	0.00	0.27
CXCL12	0.01	0.37	-0.01	0.26	0.34	-0.07	-0.05
IGF1	0.23	0.52	0.14	0.41	0.34	-0.11	0.09
ITGA11	-0.11	-0.22	-0.37	-0.11	0.20	-0.07	-0.11
IVL	-0.30	-0.41	-0.30	-0.43	-0.41	-0.50	-0.42
LEF1	-0.03	0.12	0.09	0.15	0.12	-0.04	0.07
PRB2	-0.85	-0.93	-0.93	-0.82	-0.87	-0.60	-0.78
SMR3B	-0.47	-0.13	-0.57	-0.53	-0.51	-0.64	-0.61

IVL ($l = 5$)			
Estimand	**FLEXOR**	**IC**	**IGO**
$\lambda_l^{(1)}$	-0.45 $(-0.56, -0.10)$	-0.48 $(\mathbf{-0.63, -0.07})$	-0.45 $(\mathbf{-0.63, -0.05})$
$\lambda_l^{(2)}$	-0.61 $(-0.88, -0.39)$	-0.36 $(\mathbf{-0.90, -0.23})$	-0.37 $(\mathbf{-0.90, -0.28})$
${\sigma_l}^{(1)}$	0.88 $(0.81, 1.19)$	0.85 $(\mathbf{0.73, 1.29})$	0.88 $(\mathbf{0.70, 1.23})$
${\sigma_l}^{(2)}$	0.70 $(0.34, 0.96)$	0.90 $(\mathbf{0.29, 1.04})$	0.91 $(\mathbf{0.31, 1.03})$
$\mathrm{M}_l^{(1)}$	-0.78 $(-0.96, -0.40)$	-0.78 $(\mathbf{-1.01, -0.34})$	-0.78 $(\mathbf{-0.95, -0.38})$
$\mathrm{M}_l^{(2)}$	-0.89 $(-1.13, -0.56)$	-0.72 $(\mathbf{-1.13, -0.40})$	-0.75 $(\mathbf{-1.13, -0.47})$
$\lambda_l^{(1)} - \lambda_l^{(2)}$	0.16 $(-0.08, 0.64)$	-0.12 $(\mathbf{-0.25, 0.66})$	-0.08 $(\mathbf{-0.22, 0.74})$
$\sigma_l^{(1)}/\sigma_l^{(2)}$	1.26 $(1.00, 3.21)$	0.95 $(\mathbf{0.87, 3.36})$	0.96 $(\mathbf{0.86, 3.45})$
LEF1 ($l = 6$)			
Estimand	**FLEXOR**	**IC**	**IGO**
$\lambda_l^{(1)}$	-0.02 $(-0.21, 0.21)$	-0.01 $(\mathbf{-0.28, 0.25})$	0.04 $(-0.21, 0.20)$
$\lambda_l^{(2)}$	0.33 $(-0.00, 0.68)$	0.28 $(\mathbf{-0.30, 0.68})$	0.35 $(\mathbf{-0.14, 0.74})$
${\sigma_l}^{(1)}$	0.93 $(0.77, 1.18)$	0.90 $(0.81, 1.17)$	0.93 $(0.80, 1.15)$
${\sigma_l}^{(2)}$	0.72 $(0.54, 1.11)$	0.61 $(\mathbf{0.53, 1.13})$	0.66 $(0.53, 1.07)$
$\mathrm{M}_l^{(1)}$	-0.10 $(-0.33, 0.16)$	-0.05 $(\mathbf{-0.38, 0.14})$	-0.03 $(\mathbf{-0.38, 0.16})$
$\mathrm{M}_l^{(2)}$	0.22 $(-0.09, 0.62)$	0.20 $(\mathbf{-0.22, 0.75})$	0.24 $(\mathbf{-0.17, 0.72})$
$\lambda_l^{(1)} - \lambda_l^{(2)}$	-0.35 $(-0.65, 0.07)$	-0.29 $(\mathbf{-0.79, 0.46})$	-0.31 $(\mathbf{-0.78, 0.23})$
$\sigma_l^{(1)}/\sigma_l^{(2)}$	1.29 $(0.91, 1.98)$	1.47 $(\mathbf{0.86, 1.90})$	1.41 $(\mathbf{0.85, 1.88})$
PRB2 ($l = 7$)			
Estimand	**FLEXOR**	**IC**	**IGO**
$\lambda_l^{(1)}$	-0.82 $(-0.88, -0.69)$	-0.83 $(\mathbf{-0.90, -0.66})$	-0.84 $(\mathbf{-0.89, -0.65})$
$\lambda_l^{(2)}$	-0.88 $(-0.95, -0.50)$	-0.87 $(\mathbf{-0.95, -0.04})$	-0.89 $(\mathbf{-0.95, -0.31})$
${\sigma_l}^{(1)}$	0.38 $(0.20, 0.79)$	0.41 $(\mathbf{0.21, 0.94})$	0.40 $(\mathbf{0.22, 0.95})$
${\sigma_l}^{(2)}$	0.19 $(0.00, 1.48)$	0.23 $(\mathbf{0.00, 2.08})$	0.20 $(\mathbf{0.00, 1.82})$
$\mathrm{M}_l^{(1)}$	-0.95 $(-0.95, -0.95)$	-0.95 $(-0.95, -0.95)$	-0.95 $(-0.95, -0.95)$
$\mathrm{M}_l^{(2)}$	-0.95 $(-0.95, -0.95)$	-0.95 $(-0.95, -0.95)$	-0.95 $(-0.95, -0.95)$
$\lambda_l^{(1)} - \lambda_l^{(2)}$	0.06 $(-0.30, 0.21)$	0.04 $(-0.80, 0.29)$	0.05 $(-0.50, 0.28)$
SMR3B ($l = 8$)			
Estimand	**FLEXOR**	**IC**	**IGO**
$\lambda_l^{(1)}$	-0.65 $(-0.78, -0.40)$	-0.69 $(\mathbf{-0.79, -0.19})$	-0.67 $(\mathbf{-0.77, -0.26})$
$\lambda_l^{(2)}$	0.04 $(-0.50, 0.29)$	-0.04 $(\mathbf{-0.57, 0.39})$	-0.11 $(\mathbf{-0.60, 0.41})$
${\sigma_l}^{(1)}$	0.68 $(0.36, 1.15)$	0.65 $(\mathbf{0.31, 1.48})$	0.68 $(\mathbf{0.36, 1.42})$
${\sigma_l}^{(2)}$	1.01 $(0.68, 1.26)$	1.05 $(\mathbf{0.54, 1.28})$	1.02 $(\mathbf{0.58, 1.32})$
$\mathrm{M}_l^{(1)}$	-0.94 $(-0.94, -0.94)$	-0.94 $(\mathbf{-0.94, -0.86})$	-0.94 $(-0.94, -0.94)$
$\mathrm{M}_l^{(2)}$	-0.25 $(-0.94, 0.06)$	-0.25 $(\mathbf{-0.94, 0.42})$	-0.57 $(\mathbf{-0.94, 0.26})$
$\lambda_l^{(1)} - \lambda_l^{(2)}$	-0.70 $(-0.88, -0.15)$	-0.65 $(\mathbf{-1.07, 0.07})$	-0.56 $(\mathbf{-1.11, 0.02})$
$\sigma_l^{(1)}/\sigma_l^{(2)}$	0.67 $(0.38, 1.38)$	0.61 $(\mathbf{0.36, 1.80})$	0.66 $(\mathbf{0.37, 1.70})$

Table 3: For four targeted genes, estimates and 95% bootstrap confidence levels (shown in parenthesis) of different population-level estimands of the potential outcomes of group 1 (IDC cancer subtype, indicated by superscript 1) and group 2 (ILC cancer subtype, indicated by superscript 2) with FLEXOR, IC, and IGO weights. An IC or IGO confidence interval is highlighted in bold if it is *wider* than the FLEXOR confidence interval for that estimand. For gene PRB2, $\sigma_l^{(1)}/\sigma_l^{(2)}$ was estimated to be much larger than 1 in all the bootstrap samples, and is not shown. See Section 5 for further explanation.

FLEXOR Pseudo-population								
Infiltrating Ductal Carcinoma								
	COL9A3	CXCL12	IGF1	ITGA11	IVL	LEF1	PRB2	SMR3B
COL9A3	1	$(-0.4, 0.2)$	$(-0.4, 0.1)$	$(-0.4, 0.1)$	$(-0.1, 0.4)$	$(-0.4, -0.0)$	$(0.0, 0.5)$	$(-0.1, 0.3)$
CXCL12		1	$(0.6, 0.8)$	$(0.3, 0.7)$	$(-0.4, 0.2)$	$(0.2, 0.7)$	$(-0.4, 0.1)$	$(-0.4, 0.4)$
IGF1			1	$(0.1, 0.5)$	$(-0.4, 0.1)$	$(0.2, 0.6)$	$(-0.4, 0.2)$	$(-0.3, 0.4)$
ITGA11				1	$(-0.3, 0.1)$	$(0.0, 0.5)$	$(-0.3, 0.1)$	$(-0.4, 0.1)$
IVL					1	$(-0.5, -0.1)$	$(-0.0, 0.5)$	$(-0.2, 0.4)$
LEF1						1	$(-0.3, 0.2)$	$(-0.4, 0.2)$
PRB2							1	$(-0.1, 0.1)$
SMR3B								1
Infiltrating Lobular Carcinoma								
	COL9A3	CXCL12	IGF1	ITGA11	IVL	LEF1	PRB2	SMR3B
COL9A3	1	$(-0.2, 0.6)$	$(-0.4, 0.4)$	$(-0.3, 0.5)$	$(-0.3, 0.5)$	$(-0.4, 0.3)$	$(-0.1, 0.3)$	$(-0.3, 0.6)$
CXCL12		1	$(0.4, 0.9)$	$(-0.1, 0.7)$	$(-0.4, 0.5)$	$(-0.4, 0.6)$	$(-0.8, 0.5)$	$(-0.3, 0.5)$
IGF1			1	$(-0.4, 0.5)$	$(-0.4, 0.3)$	$(-0.4, 0.5)$	$(-0.7, 0.3)$	$(-0.4, 0.5)$
ITGA11				1	$(-0.4, 0.5)$	$(-0.4, 0.5)$	$(-0.5, 0.4)$	$(-0.5, 0.3)$
IVL					1	$(-0.6, 0.1)$	$(-0.2, 0.7)$	$(-0.3, 0.5)$
LEF1						1	$(-0.7, 0.4)$	$(-0.4, 0.3)$
PRB2							1	$(-0.2, 0.5)$
SMR3B								1

Table 4: For FLEXOR weights, 95% confidence intervals of the mRNA expression level correlations between gene pairs for the breast cancer subtypes as the groups.

IC Pseudo-population

Infiltrating Ductal Carcinoma

	COL9A3	CXCL12	IGF1	ITGA11	IVL	LEF1	PRB2	SMR3B
COL9A3	1	$(-0.4, 0.2)$	$(-0.4, 0.2)$	$(-0.4, 0.2)$	$(-0.1, 0.4)$	$(-0.4, 0.0)$	$(0.1, 0.6)$	$(-0.1, 0.4)$
CXCL12		1	$(0.6, 0.9)$	$(0.2, 0.7)$	$(-0.4, 0.2)$	$(0.2, 0.7)$	$(-0.5, 0.2)$	$(-0.4, 0.3)$
IGF1			1	$(0.1, 0.5)$	$(-0.5, 0.2)$	$(0.1, 0.6)$	$(-0.5, 0.2)$	$(-0.3, 0.4)$
ITGA11				1	$(-0.3, 0.2)$	$(-0.1, 0.5)$	$(-0.3, 0.1)$	$(-0.4, 0.1)$
IVL					1	$(-0.5, 0.0)$	$(-0.1, 0.6)$	$(-0.2, 0.4)$
LEF1						1	$(-0.4, 0.2)$	$(-0.4, 0.2)$
PRB2							1	$(-0.1, 0.1)$
SMR3B								1

Infiltrating Lobular Carcinoma

	COL9A3	CXCL12	IGF1	ITGA11	IVL	LEF1	PRB2	SMR3B
COL9A3	1	$(-0.5, 0.6)$	$(-0.5, 0.6)$	$(-0.4, 0.6)$	$(-0.4, 0.5)$	$(-0.6, 0.3)$	$(-0.1, 0.4)$	$(-0.3, 0.5)$
CXCL12		1	$(0.3, 0.9)$	$(-0.3, 0.7)$	$(-0.6, 0.6)$	$(-0.7, 0.8)$	$(-0.9, 0.3)$	$(-0.4, 0.6)$
IGF1			1	$(-0.5, 0.7)$	$(-0.7, 0.4)$	$(-0.5, 0.7)$	$(-0.9, 0.2)$	$(-0.4, 0.5)$
ITGA11				1	$(-0.5, 0.7)$	$(-0.6, 0.7)$	$(-0.7, 0.3)$	$(-0.6, 0.3)$
IVL					1	$(-0.7, 0.1)$	$(-0.3, 0.8)$	$(-0.4, 0.6)$
LEF1						1	$(-0.8, 0.3)$	$(-0.6, 0.3)$
PRB2							1	$(-0.3, 0.6)$
SMR3B								1

Table 5: For IC weights, 95% confidence intervals of the mRNA expression level correlations between gene pairs for the breast cancer subtypes as the groups.

IGO Pseudo-population								
Infiltrating Ductal Carcinoma								
	COL9A3	CXCL12	IGF1	ITGA11	IVL	LEF1	PRB2	SMR3B
COL9A3	1	$(-0.3, 0.2)$	$(-0.4, 0.1)$	$(-0.3, 0.2)$	$(-0.1, 0.4)$	$(-0.4, 0.1)$	$(0.1, 0.5)$	$(-0.1, 0.4)$
CXCL12		1	$(0.6, 0.8)$	$(0.2, 0.7)$	$(-0.4, 0.1)$	$(0.2, 0.6)$	$(-0.5, 0.2)$	$(-0.4, 0.3)$
IGF1			1	$(0.1, 0.5)$	$(-0.5, 0.2)$	$(0.1, 0.6)$	$(-0.4, 0.2)$	$(-0.3, 0.3)$
ITGA11				1	$(-0.4, 0.2)$	$(-0.0, 0.5)$	$(-0.3, 0.1)$	$(-0.4, 0.1)$
IVL					1	$(-0.5, 0.0)$	$(-0.1, 0.5)$	$(-0.2, 0.4)$
LEF1						1	$(-0.4, 0.2)$	$(-0.4, 0.2)$
PRB2							1	$(-0.1, 0.1)$
SMR3B								1
Infiltrating Lobular Carcinoma								
	COL9A3	CXCL12	IGF1	ITGA11	IVL	LEF1	PRB2	SMR3B
COL9A3	1	$(-0.4, 0.6)$	$(-0.4, 0.5)$	$(-0.5, 0.6)$	$(-0.3, 0.5)$	$(-0.6, 0.4)$	$(-0.1, 0.4)$	$(-0.3, 0.5)$
CXCL12		1	$(0.4, 0.9)$	$(-0.2, 0.7)$	$(-0.6, 0.5)$	$(-0.5, 0.7)$	$(-0.8, 0.5)$	$(-0.6, 0.4)$
IGF1			1	$(-0.4, 0.6)$	$(-0.6, 0.4)$	$(-0.5, 0.6)$	$(-0.8, 0.3)$	$(-0.5, 0.4)$
ITGA11				1	$(-0.5, 0.6)$	$(-0.4, 0.5)$	$(-0.6, 0.4)$	$(-0.6, 0.3)$
IVL					1	$(-0.6, 0.0)$	$(-0.2, 0.7)$	$(-0.4, 0.5)$
LEF1						1	$(-0.8, 0.3)$	$(-0.5, 0.2)$
PRB2							1	$(-0.3, 0.6)$
SMR3B								1

Table 6: For IGO weights, 95% confidence intervals of the mRNA expression level correlations between gene pairs for the breast cancer subtypes as the groups.