Results so Far(Univariate Analysis)

Olusoji Oluwafemi Daniel (1541893)

Hasselt University, Belgium

June 3, 2017

Contents

Background & Objectives

Background

- The human microbiome is made up of trillions of symbiotic bacteria cells in humans(Ursell, et al, 2013).
- Although their functions is not yet fully understood, they are associated with nutrition, metabolism, immune function and human physiology(Bull & Plummer, 2013).
- The part of the human microbiome that reside in the gastrointestinal tract(GUT) are even associated with more functions than others(Bull & Plummer, 2013).
- Changes in the GUT microbiome have been associated with; obesity, lipid accumulation and metabolic disorder(Boulange, et al, 2016).
- However, the exact causal relationships between this microbiome ecosystem and health conditions is not yet fully understood due to the complex pathology involved(Boulange, et al, 2016).
- These established associations are however enough to trigger scientists and pharmaceutical establishments to develop treatments that target these microbiome ecosystem, to trigger a benefitial treatment effect in patients with associated health disorders.
- But if this is to be achieved, elements of the GUT microbiome must be tied to health benefits in these patients, hence they must be good surrogates for already established measures of treatment effect in these patients(e.g., insulin sensitivity).

Objectives

The main objective of this thesis is to discover bacterias in gut microbiome that can serve as surrogate for insulin sensitivity in patients with metabolic syndrome. Specifically;

- it is of interest to examine each of 30 candidates as a potential surrogate for insulin sensitivity,
- Identify a combination of these 30 that can be a surrogate for insulin sensitivity.

Notations for Response, Surrogates and Treatment

- $\mathbf{T_i}$ = difference in insulin sensitivity, which is the true endpoint, $i = 1, \ldots, 16$
- $\mathbf{S_{ij}}$ = surrogates (difference in bacteria counts), $j = 1, \ldots, 30$.
- $\mathbf{X_i}$ = Treatment applied. own_feces(control), donor_feces(experimental).
- C=control treatment and E=Experimental treatment

Joint Modelling Approach(JMA)

- The joint model framework to surrogate evaluation (Molenberghs, Burzykowski, Renard and Geys, 2000, Burzykowski, Molenberghs, and Buyse, 2005, Buyse, et al, 2015, Perualia-Tan, et al, 2016) can be extended to the following surrogate specific joint model;

$$\begin{pmatrix} S_{ij} \\ T_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_S + \alpha_j X_i \\ \mu_T + \beta X_i \end{pmatrix}, \Sigma_j\right], \Sigma_j = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ \sigma_{ST} & \sigma_{TT} \end{pmatrix}$$

- α_j is the effect of treatment on surrogate j
- β is the effect of treatment on the true endpoint
- $\rho_{T,S|X} = \frac{\sigma_{ST}}{\sqrt{\sigma_{SS}\sigma_{TT}}}$. This is the adjusted association (Burzykowski, Molenberghs, and Buyse, 2005, Perualia-Tan, et al, 2016).
- To test hypothesis about $\rho_{T,S|X}$, a reduced form of the joint model above can be fitted. Specificaly, the joint model of interest is;

$$\begin{pmatrix} S_{ij} \\ T_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_S + \alpha_j X_i \\ \mu_T + \beta X_i \end{pmatrix}, \Sigma_j\right], \Sigma_j = \begin{pmatrix} \sigma_{SS} & 0 \\ 0 & \sigma_{TT} \end{pmatrix}$$

- a comparison of both joint models with likelihood ratio tests will test the null hypothesis, $H_0 : \rho_{T,S|X} = 0$(Perualia-Tan, et al, 2016)

Causal Inference Approach(CIA)

- The CIA approach to evaluation of surrogates came up as a result of the difficulty associated with establishing surrogacy in single trials(Van der Elst, Molenberghs and Alonso, 2015).
- The approach (also termed counter-factual approach to surrogate endpoint evaluation), involves assuming the patient has four potential outcomes$Y = (T_0, T_1, S_0, S_1) \sim N(\mu, \Sigma), \Sigma =$

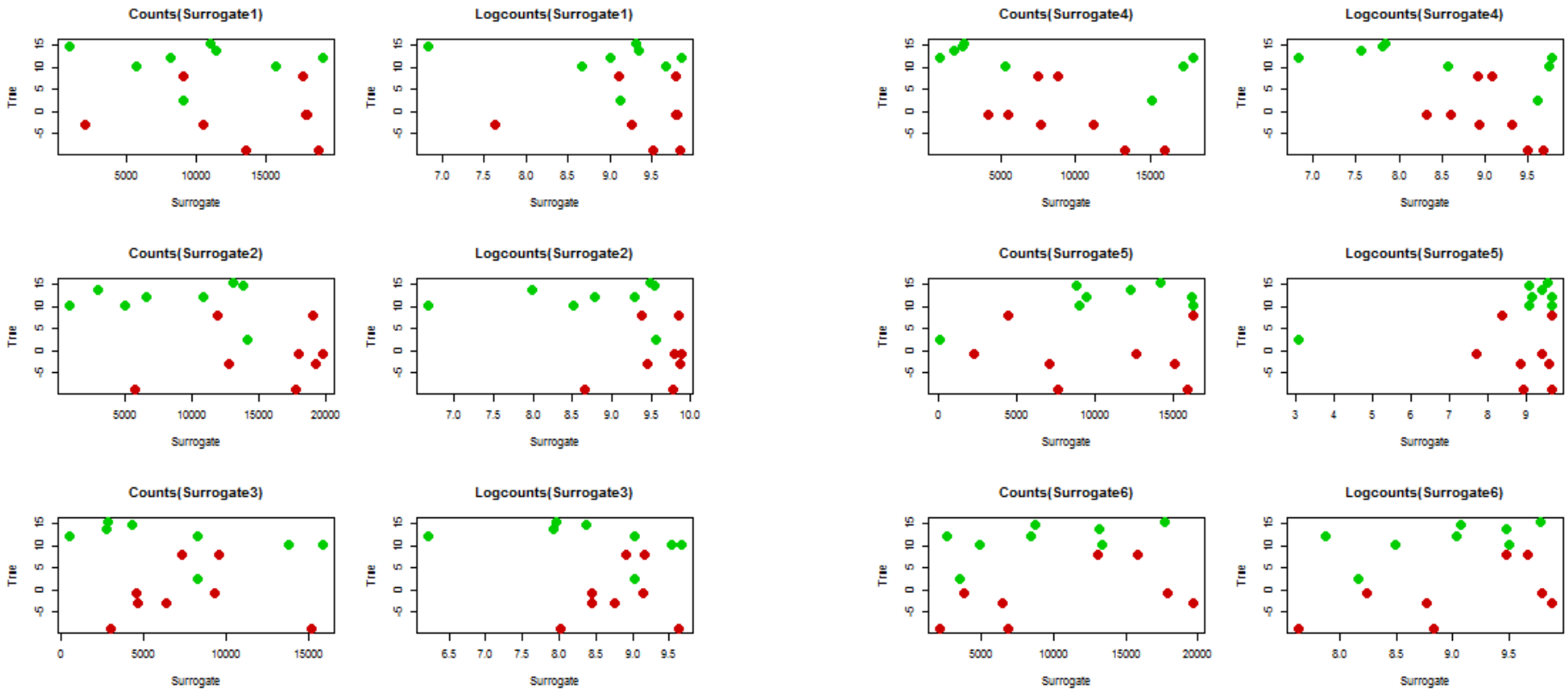
$$\left(\begin{array}{cc|cc} \sigma_{T_0T_0} & \sigma_{T_0T_1} & \sigma_{T_0S_0} & \sigma_{T_0S_1} \\ \sigma_{T_0T_1} & \sigma_{T_1T_1} & \sigma_{T_1S_0} & \sigma_{T_1S_1} \\ \hline \sigma_{T_0S_0} & \sigma_{T_1S_0} & \sigma_{S_0S_0} & \sigma_{S_0S_1} \\ \sigma_{T_0S_1} & \sigma_{T_1S_1} & \sigma_{S_0S_1} & \sigma_{S_1S_1} \end{array}\right).$$

- Alonso et al, 2015 used the individual causal association ($ICA = \frac{\sqrt{\sigma_{T_0T_0}\sigma_{S_0S_0}}\rho_{T_0S_0} + \sqrt{\sigma_{T_1T_1}\sigma_{S_1S_1}}\rho_{T_1S_1} - \sqrt{\sigma_{T_1T_1}\sigma_{S_0S_0}}\rho_{T_1S_0} - \sqrt{\sigma_{T_0T_0}\sigma_{S_1S_1}}\rho_{T_0S_1}}{\sqrt{(\sigma_{T_0T_0}+\sigma_{T_1T_1}-2\sqrt{\sigma_{T_0T_0}\sigma_{T_1T_1}}\rho_{T_0T_1})(\sigma_{S_0S_0}+\sigma_{S_1S_1}-2\sqrt{\sigma_{S_0S_0}\sigma_{S_1S_1}}\rho_{S_0S_1})}}$) to assess surrogacy in single trials.
- The quantities $\Delta_T = T_1 - T_0, \Delta_S = S_1 - S_0$ are called the individual causal effects. Since these causal effects are simply a linear combination of the potential outcomes, then

$$\Delta_T, \Delta_S = \begin{pmatrix} T_1 - T_0 \\ S_1 - S_0 \end{pmatrix} \sim N(\mu_\Delta, \Sigma_\Delta), \Sigma_\Delta = A\Sigma A'$$

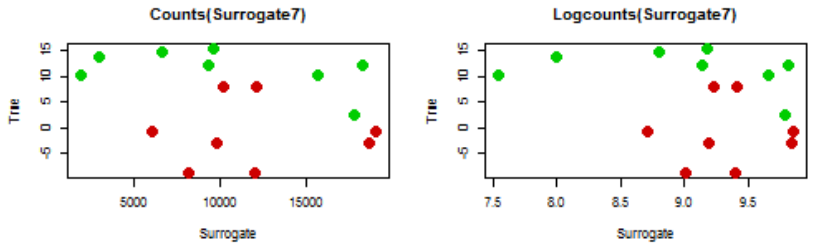
- This approach is involved, since half of the needed data is missing.

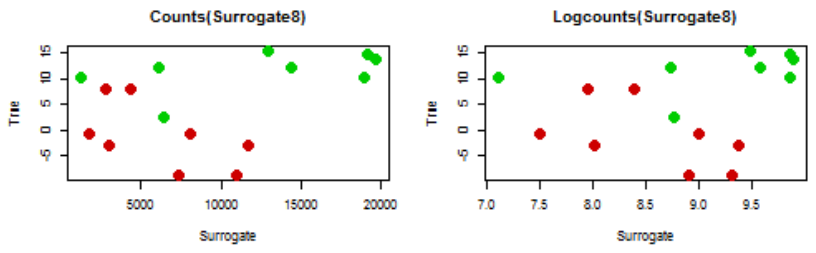
Exploring Adjusted Association

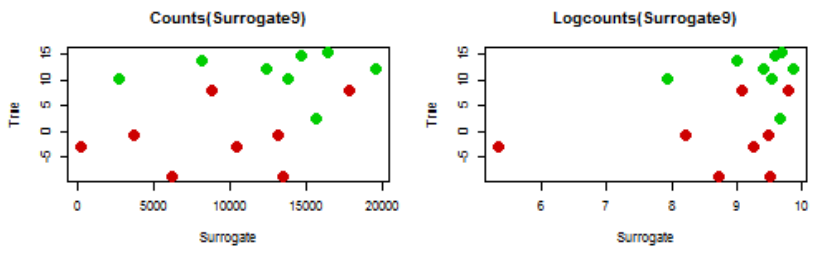


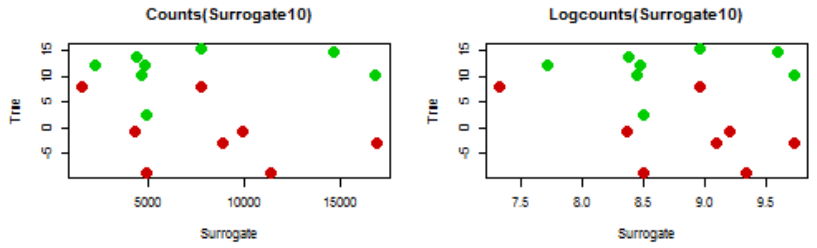
- red = control treatment, green = experimental treatment
- y axis shows treatment effect on T
- x axis shows treatment effect on S
- the arrangement of the scattered points gives an idea of the adjusted association.

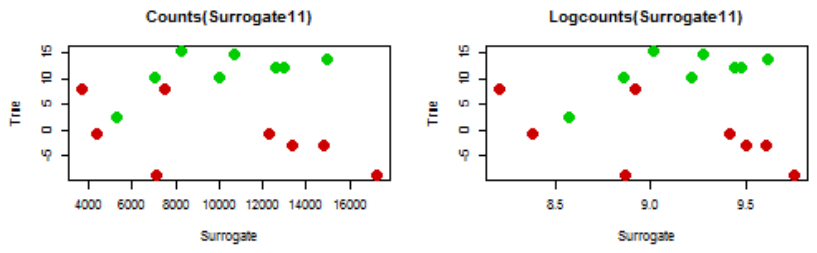
Exploring Adjusted Association(cont:)

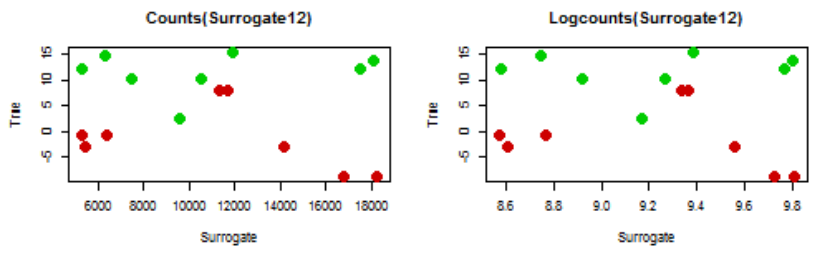




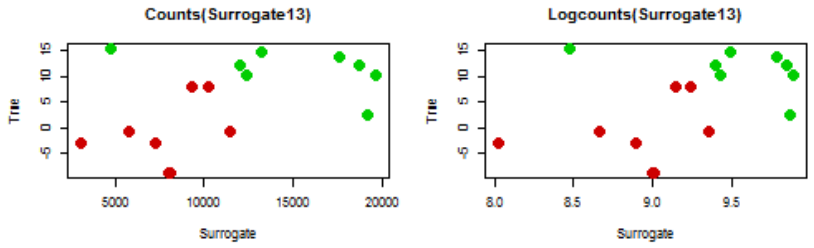


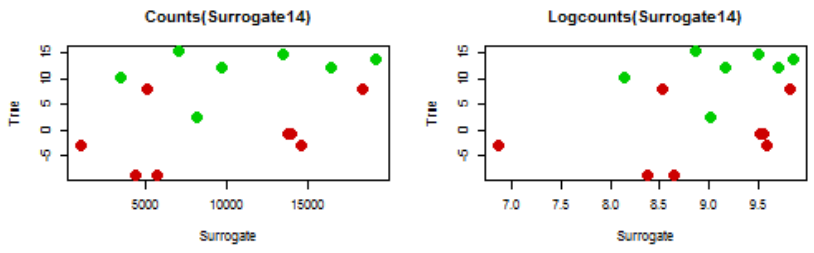


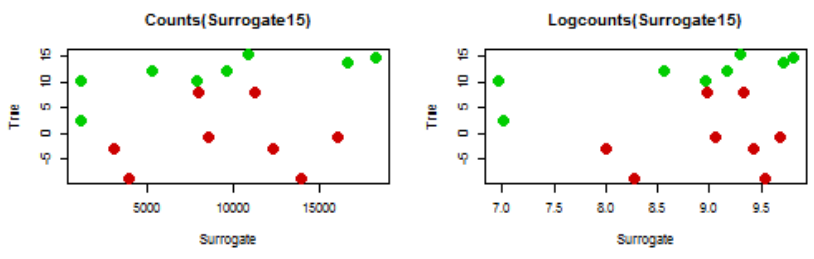


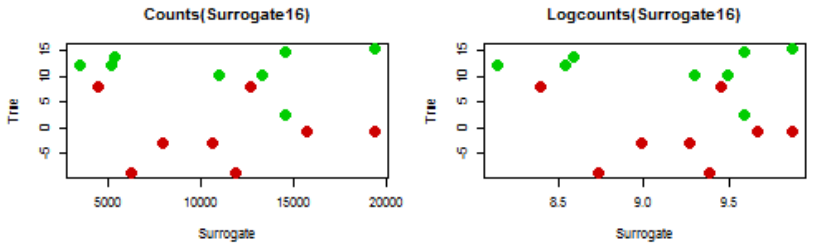


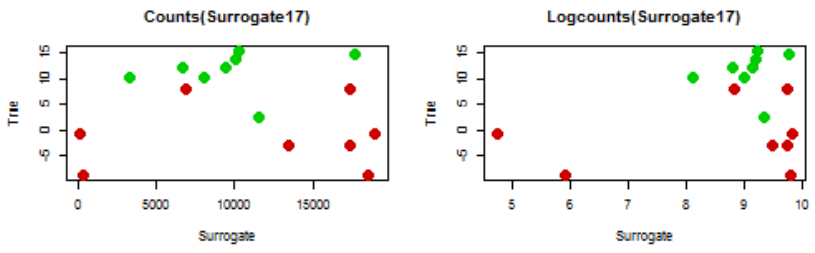
Exploring Adjusted Association(cont:)

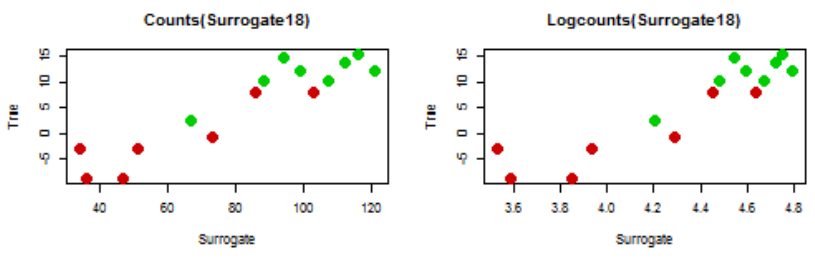




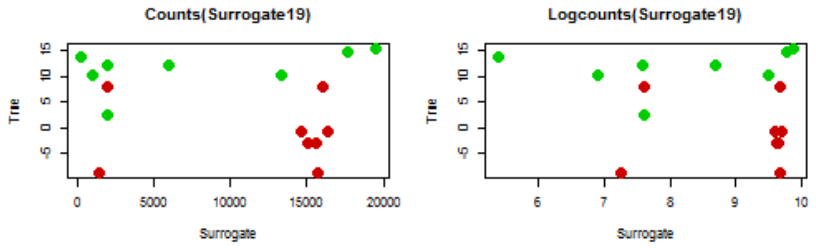


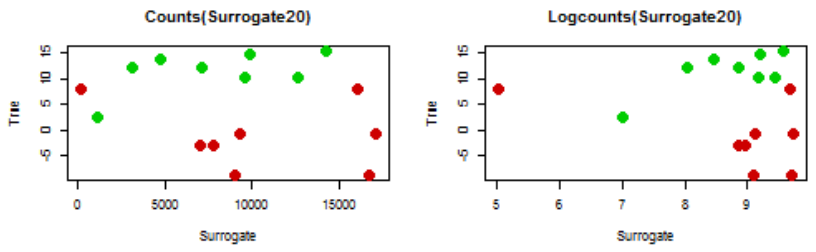


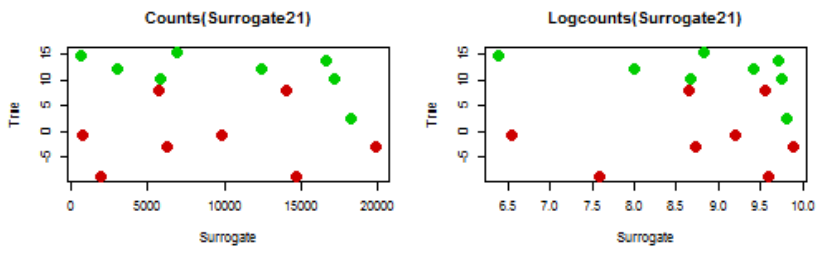


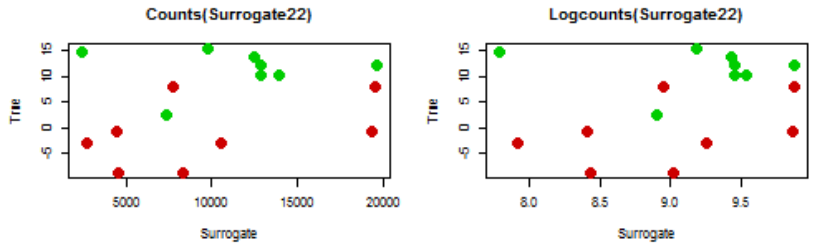


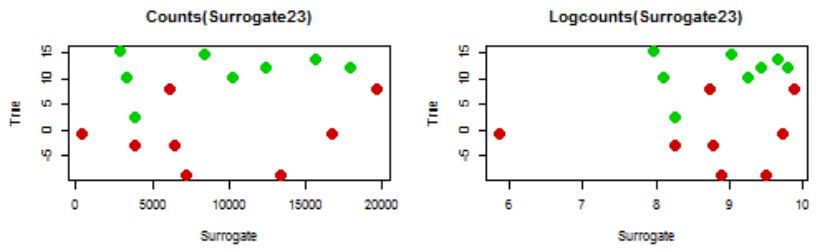
Exploring Adjusted Association(cont:)

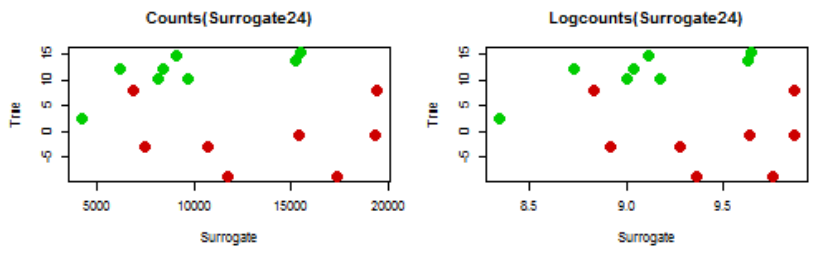




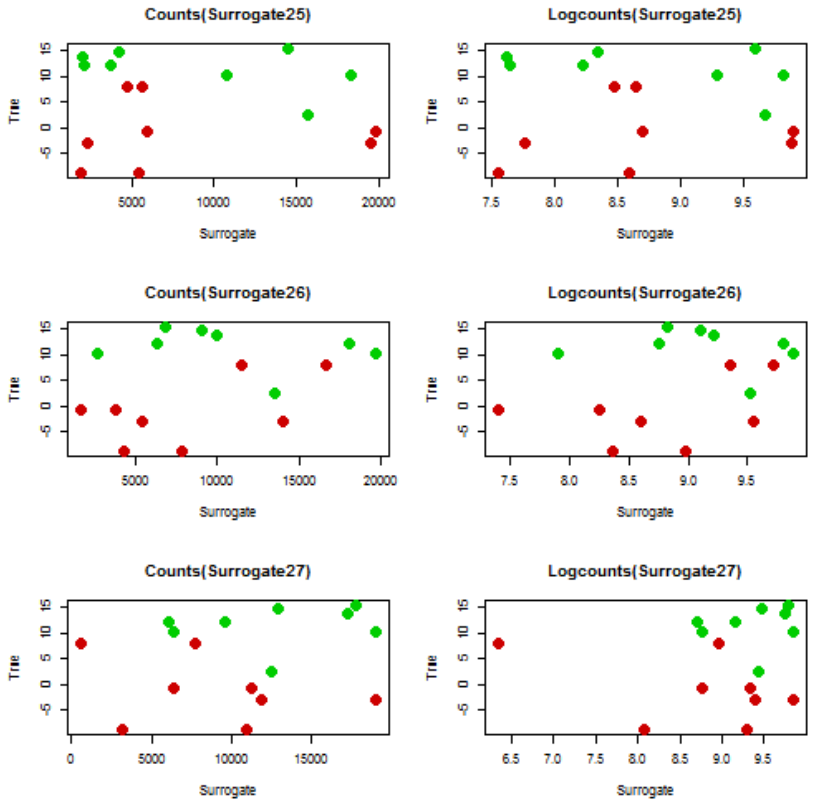


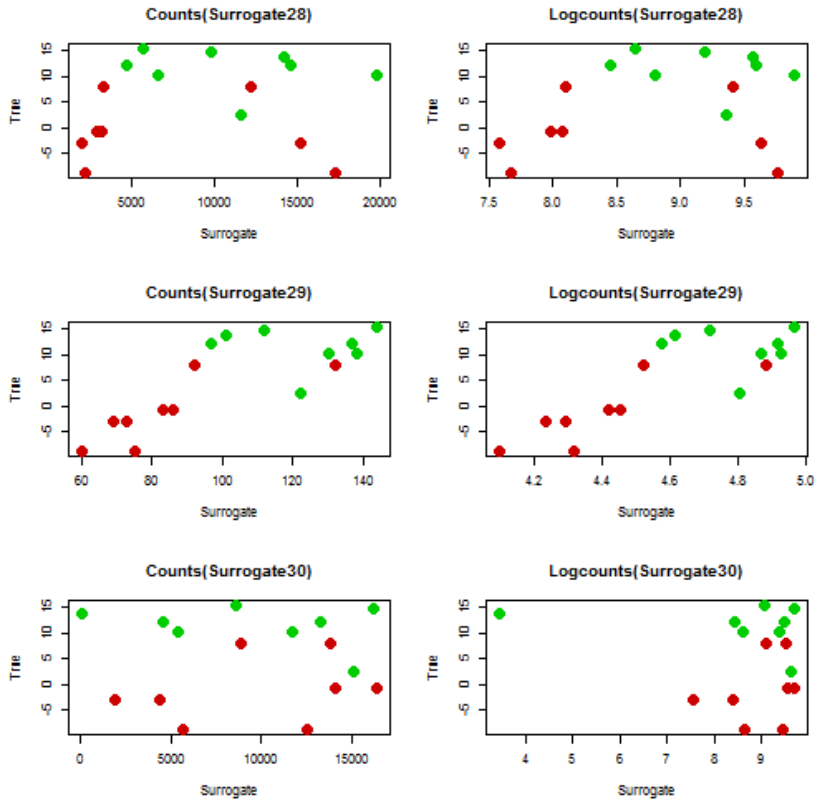






Exploring Adjusted Association(cont:)





Treatment Effect on Insulin Sensitivity(T)

A paired t-test and Wilcoxon test(due to the sample size) is used to test for treatmment effect on T. The null hypothesis of interest is;

$$H_0 : \mu_{T_c} = \mu_{T_E}$$

t-test(treatment effect=difference in means)

Treatment Effect	95%CI
1257	6.68, 18.47

Wilcoxon-test(treatment effect=difference in Ranks)

Treatment Effect	95%CI
1323	5.45, 18.96

Both tests showed a treatment effect in favor of the experimental treatment.

Treatment Effect on Surrogates

Only the log of counts will be used from here onward.
The null hypothesis of interest remains;

$$H_0 : \mu_{S_C} = \mu_{S_E}$$

for each of the 30 surrogates.

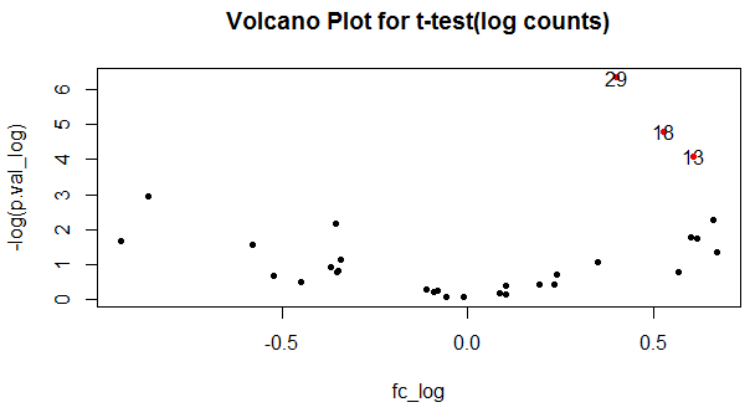


Figure 1: t-tests

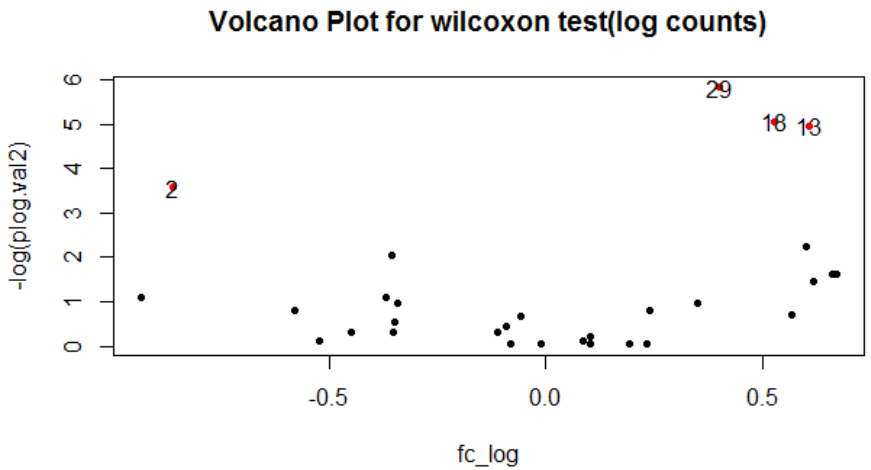


Figure 2: Wilcoxon-tests

Adjusted Treatment Effect on Surrogates

S	rawp	BH	fc
29	0.00	0.05	-3272.38
18	0.01	0.12	-7137.25
13	0.02	0.17	-417.62
2	0.05	0.39	-1358.12
28	0.10	0.57	632.38
24	0.11	0.57	-1647.00
8	0.17	0.63	-1749.12
27	0.18	0.63	6084.88
19	0.19	0.63	3720.62
4	0.21	0.64	-674.75
9	0.26	0.70	186.50
7	0.32	0.80	-328.88
26	0.34	0.80	6791.50
1	0.40	0.81	524.62
3	0.45	0.81	-788.38
15	0.45	0.81	-295.00
17	0.47	0.81	-1976.50
22	0.49	0.81	37.62
30	0.52	0.82	-4382.62
5	0.61	0.88	-2595.75
14	0.66	0.88	943.00
23	0.66	0.88	1754.12
11	0.68	0.88	106.88
6	0.77	0.92	-3958.62
16	0.78	0.92	741.88
10	0.80	0.92	2626.62
25	0.85	0.92	3796.00
21	0.86	0.92	3601.38
20	0.93	0.96	38.88
12	0.96	0.96	-346.50

Table 1: t-tests

Adjusted Treatment Effect on Surrogates

S	rawp	BH	fc
29	0.00	0.07	-3272.38
18	0.01	0.07	-7137.25
13	0.01	0.07	-417.62
2	0.03	0.21	-1358.12
8	0.10	0.63	632.38
24	0.13	0.65	-1647.00
9	0.19	0.73	-1749.12
28	0.19	0.73	6084.88
27	0.23	0.78	3720.62
1	0.33	0.88	-674.75
19	0.33	0.88	186.50
7	0.38	0.88	-328.88
26	0.38	0.88	6791.50
4	0.44	0.88	524.62
22	0.44	0.88	-788.38
17	0.49	0.89	-295.00
20	0.51	0.89	-1976.50
3	0.57	0.96	37.62
10	0.65	0.96	-4382.62
5	0.72	0.96	-2595.75
6	0.72	0.96	943.00
15	0.72	0.96	1754.12
21	0.80	0.96	106.88
25	0.88	0.96	-3958.62
30	0.88	0.96	741.88
11	0.96	0.96	2626.62
12	0.96	0.96	3796.00
14	0.96	0.96	3601.38
16	0.96	0.96	38.88
23	0.96	0.96	-346.50

Table 2: W-test

Results from Joint Models

S_i	ρ_{UN}	ρ_{ADj}	LCI	UCI	$pval_\rho$	$pval-adj_\rho$	α_j	$pval_{\alpha_j}$	$pval-adj_{\alpha_j}$
18	0.90	0.83	0.56	0.94	0.00	0.00	0.52	0.00	0.04
29	0.82	0.57	0.11	0.83	0.01	0.13	0.40	0.00	0.01
6	0.29	0.56	0.10	0.83	0.01	0.13	-0.11	0.77	0.92
15	0.12	0.46	-0.05	0.78	0.05	0.36	-0.35	0.44	0.80
5	0.14	0.41	-0.11	0.75	0.09	0.44	-0.45	0.60	0.88
14	0.29	0.32	-0.21	0.70	0.19	0.74	0.19	0.65	0.88
24	-0.18	0.25	-0.28	0.67	0.30	0.82	-0.36	0.10	0.52
22	0.29	0.23	-0.30	0.65	0.35	0.82	0.24	0.48	0.80
17	0.27	0.19	-0.34	0.63	0.44	0.82	0.56	0.45	0.80
26	0.31	0.19	-0.34	0.62	0.45	0.82	0.35	0.34	0.78
9	0.33	0.15	-0.37	0.60	0.55	0.82	0.67	0.24	0.65
23	0.17	0.12	-0.40	0.58	0.62	0.82	0.23	0.66	0.88
2	-0.34	0.10	-0.42	0.57	0.69	0.82	-0.86	0.03	0.26
19	-0.22	0.10	-0.42	0.56	0.70	0.82	-0.93	0.17	0.58
13	0.44	-0.04	-0.52	0.47	0.88	0.88	0.61	0.01	0.11
28	0.31	-0.05	-0.53	0.46	0.84	0.87	0.66	0.08	0.51
30	-0.18	-0.06	-0.54	0.45	0.80	0.86	-0.52	0.51	0.80
25	-0.01	-0.08	-0.55	0.43	0.75	0.84	0.09	0.85	0.92
1	-0.23	-0.09	-0.56	0.42	0.71	0.82	-0.37	0.40	0.80
8	0.22	-0.10	-0.56	0.42	0.70	0.82	0.60	0.16	0.58
21	-0.03	-0.11	-0.58	0.41	0.65	0.82	0.10	0.86	0.92
16	-0.13	-0.12	-0.58	0.40	0.64	0.82	-0.08	0.77	0.92
3	-0.23	-0.12	-0.58	0.40	0.64	0.82	-0.35	0.43	0.80
7	-0.29	-0.13	-0.59	0.39	0.59	0.82	-0.34	0.31	0.77
12	-0.10	-0.14	-0.59	0.38	0.58	0.82	-0.01	0.96	0.96
20	-0.14	-0.19	-0.63	0.34	0.45	0.82	-0.06	0.93	0.96
11	-0.05	-0.22	-0.64	0.31	0.38	0.82	0.10	0.67	0.88
10	-0.20	-0.24	-0.66	0.29	0.34	0.82	-0.09	0.80	0.92
27	0.10	-0.31	-0.70	0.22	0.20	0.74	0.62	0.16	0.58
4	-0.52	-0.44	-0.77	0.07	0.06	0.36	-0.58	0.19	0.58

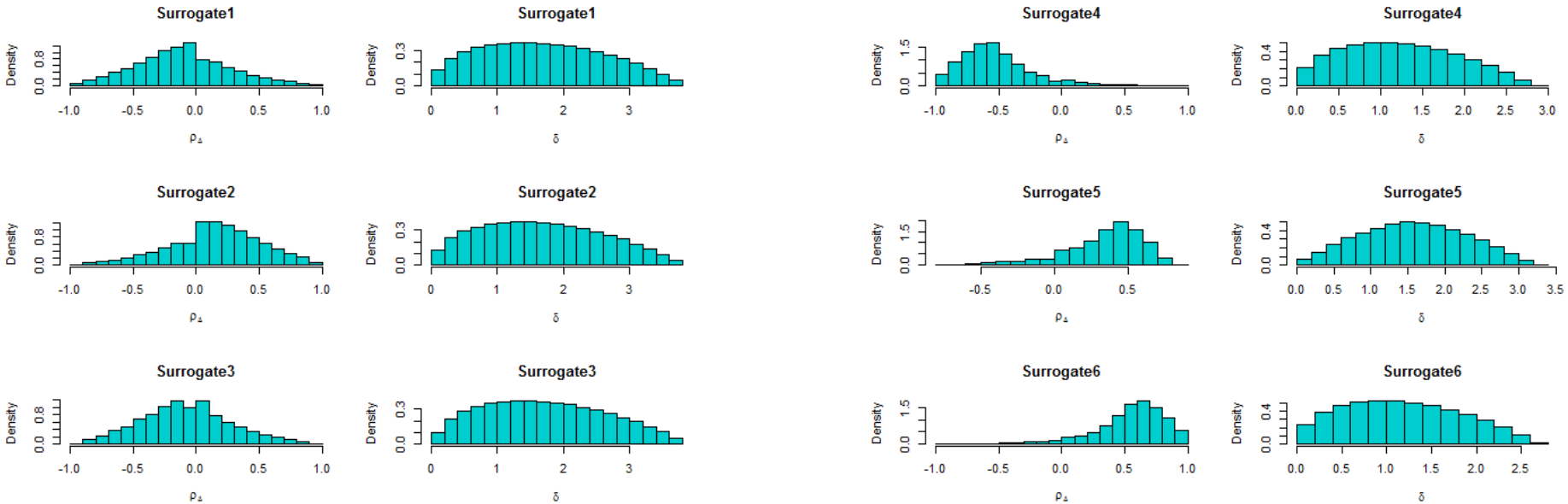
- pval=p-value, pval-adj=adjusted pvalue using BH, LCI-Lower Confidence Interval and UCI=Upper Confidence Interval for ρ and results are ordered according to adjusted association.

Causal Inference Results

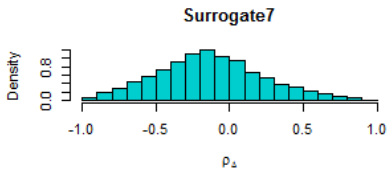
S	ρ_0	LCI_0	UCI_0	ρ_1	LCI_1	UCI_1	ICA	LRan	URan
18	0.84	0.60	0.94	0.81	0.52	0.93	0.91	-0.76	1.00
6	0.57	0.11	0.83	0.56	0.10	0.83	0.84	-0.96	1.00
15	0.20	-0.33	0.63	0.79	0.49	0.93	0.83	-0.91	0.93
17	0.21	-0.32	0.64	0.14	-0.39	0.59	0.74	-1.00	1.00
19	0.00	-0.49	0.50	0.21	-0.32	0.64	0.70	-0.99	0.99
24	-0.10	-0.57	0.41	0.81	0.52	0.93	0.69	-0.85	0.86
29	0.85	0.60	0.95	-0.11	-0.57	0.41	0.67	-0.81	0.84
14	0.32	-0.21	0.70	0.32	-0.21	0.70	0.63	-0.99	1.00
5	-0.16	-0.60	0.37	0.88	0.67	0.96	0.60	-0.73	0.80
22	0.43	-0.09	0.76	-0.10	-0.57	0.42	0.58	-0.96	0.95
2	0.32	-0.21	0.70	-0.01	-0.50	0.49	0.57	-0.98	0.98
20	-0.51	-0.80	-0.02	0.73	0.37	0.90	0.53	-0.75	0.77
23	0.06	-0.45	0.54	0.30	-0.23	0.69	0.52	-0.99	0.98
8	-0.52	-0.81	-0.03	0.37	-0.15	0.73	0.52	-0.87	0.87
26	0.39	-0.14	0.74	-0.19	-0.63	0.34	0.51	-0.95	0.95
11	-0.57	-0.83	-0.10	0.70	0.32	0.89	0.48	-0.75	0.74
9	0.19	-0.34	0.63	0.01	-0.49	0.50	0.48	-0.99	0.99
3	0.20	-0.33	0.63	-0.38	-0.74	0.14	0.34	-0.94	0.94
30	0.28	-0.25	0.68	-0.28	-0.68	0.25	0.33	-0.95	0.95
28	0.02	-0.48	0.51	-0.23	-0.65	0.30	0.31	-0.99	0.99
1	0.01	-0.49	0.50	-0.24	-0.66	0.29	0.30	-0.99	0.98
12	-0.27	-0.67	0.26	0.09	-0.43	0.56	0.28	-0.97	0.98
13	0.29	-0.24	0.69	-0.50	-0.80	-0.00	0.17	-0.91	0.90
7	0.06	-0.45	0.54	-0.32	-0.70	0.21	0.14	-0.98	0.97
16	-0.10	-0.57	0.42	-0.15	-0.60	0.37	0.13	-0.99	0.99
25	0.18	-0.34	0.62	-0.47	-0.78	0.04	0.05	-0.93	0.93
27	-0.44	-0.77	0.07	0.16	-0.37	0.61	0.05	-0.95	0.95
10	-0.48	-0.79	0.02	0.18	-0.35	0.62	0.05	-0.93	0.94
21	0.12	-0.40	0.58	-0.47	-0.78	0.03	0.03	-0.95	0.94
4	-0.49	-0.79	0.01	-0.59	-0.84	-0.13	-0.79	-1.00	0.97

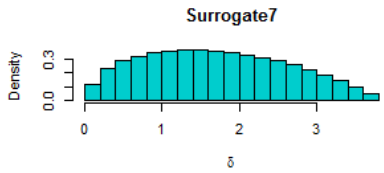
Table 3: Ranked by ICA values

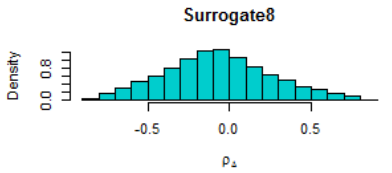
Distribution of ICA

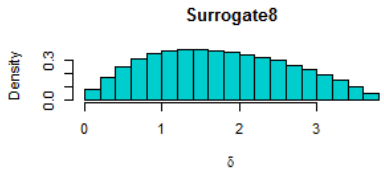


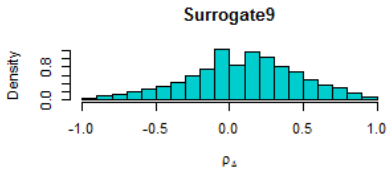
Distribution of ICA(cont:)

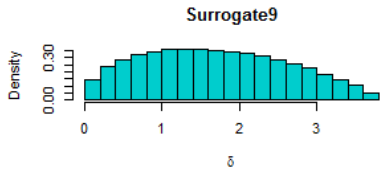


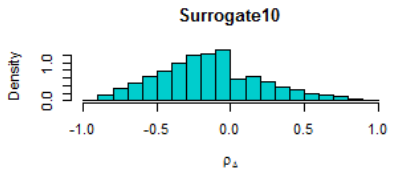


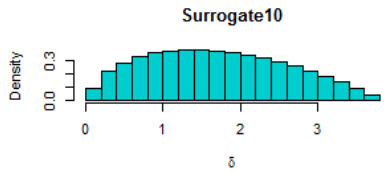


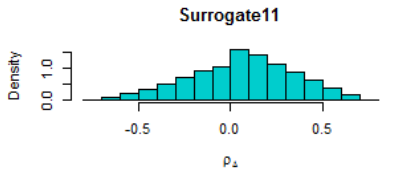


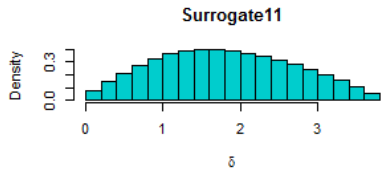


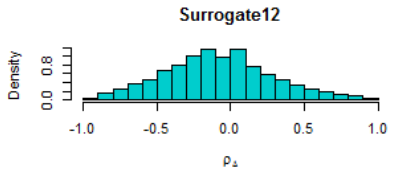


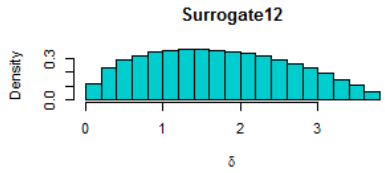




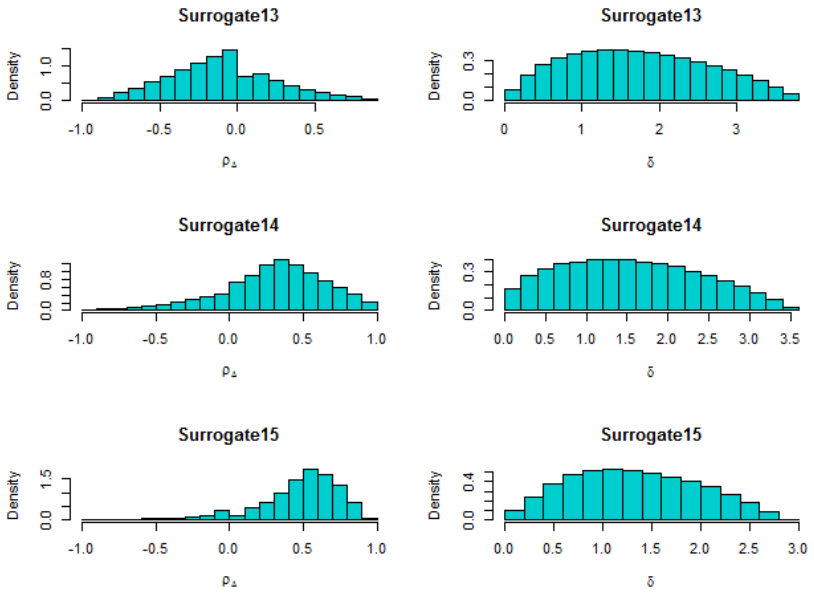


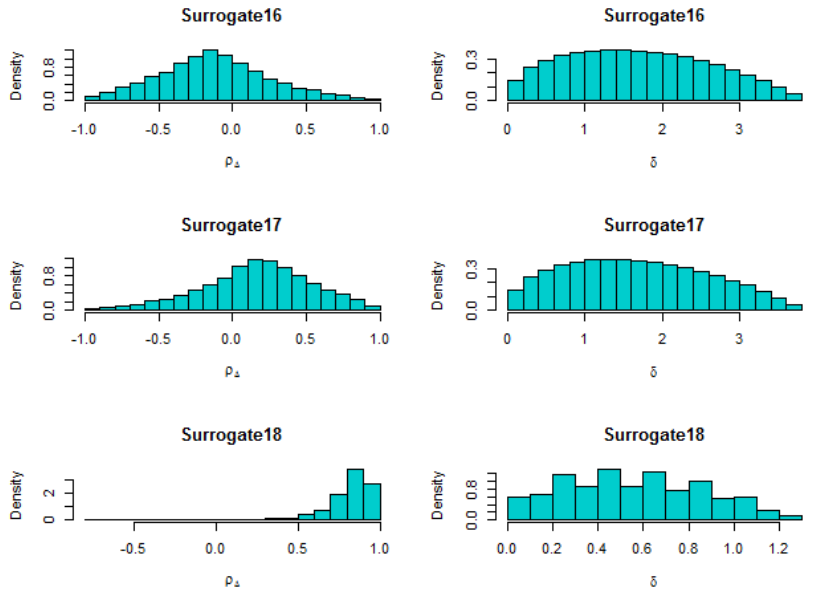




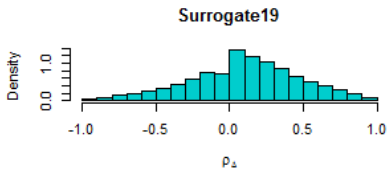


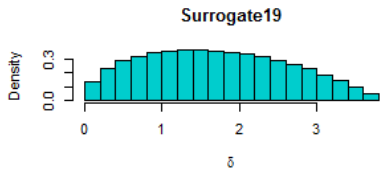
Distribution of ICA(cont:)

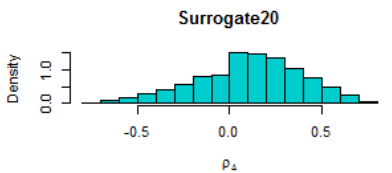


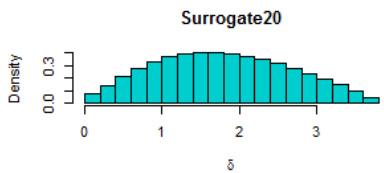


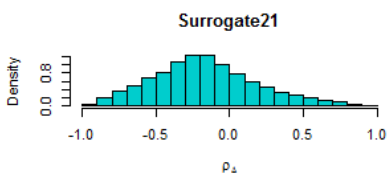
Distribution of ICA(cont:)

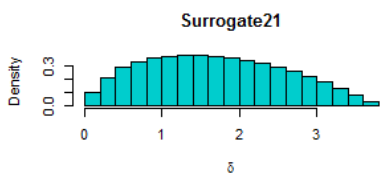


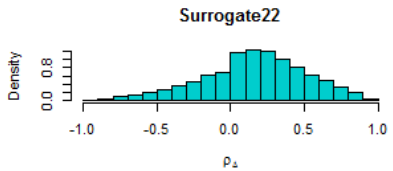


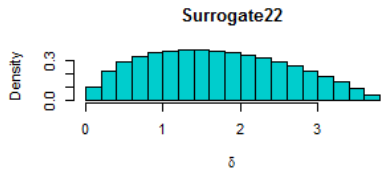


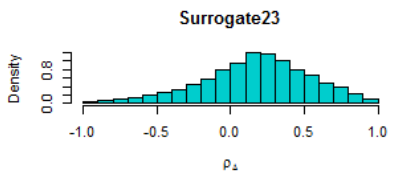


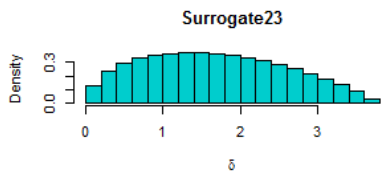


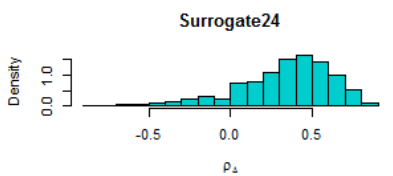


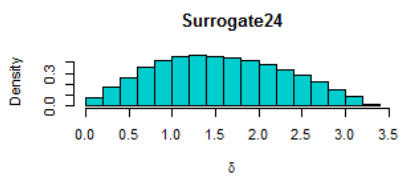




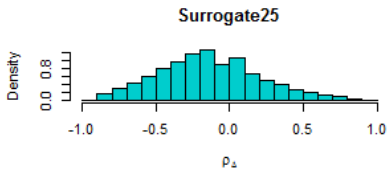


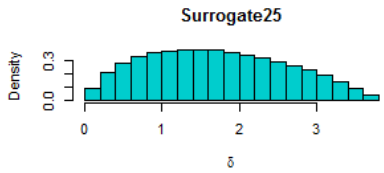


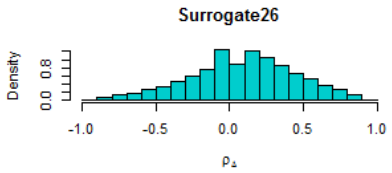


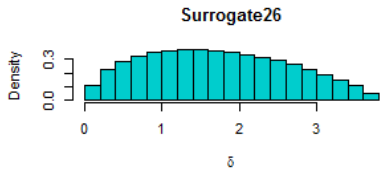


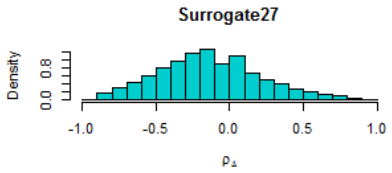
Distribution of ICA(cont:)

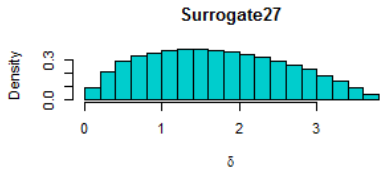


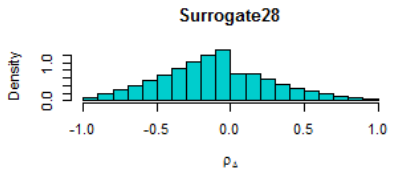


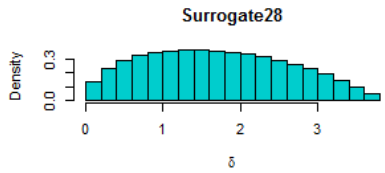


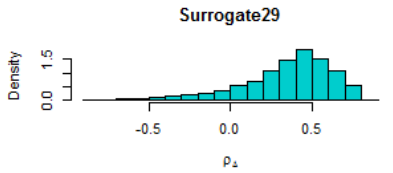


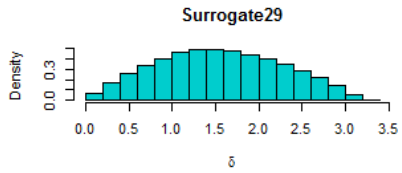


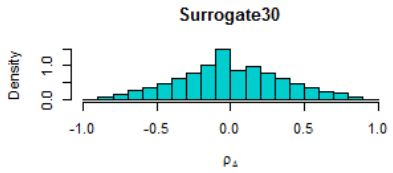


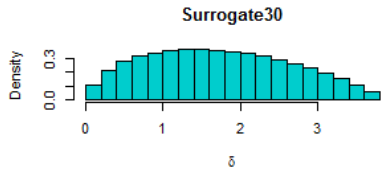












What We Learn from The Results

- The exploration revealed the possibility of high adjusted association between surrogate 29, 18 and T.
- The joint model confirmed the suspicion of high adjusted association($\rho_{T,S|X} = 0.83$) between 18 and T. The obtained ICA for this surrogate is also high ($\rho_{\Delta} = 0.91$) but the range of obtained ICA values is quite large (-0.76, 1) but in most cases, these ICA values are larger than 0.7(see ICA distribution for surrogate 18).
- The adjusted association for surrogate 29 was a moderate association ($\rho_{T,S|X} = 0.57$). Results from the causal inference framework showed that the relationship between T and surrogate 29 is opposite in the treatment groups ($\rho_0 = 0.85, \rho_1 = -0.11$). This is potentially responsible for the moderate adjusted association,
- Studying the top 10 surrogates in each approach showed that only surrogate 19 made it to the top 10 in CIA that was not present in the JMA, while only surrogate 26 made it in JMA and is not present in CIA. Notably, surrogate 26 had opposite relationship between T and S in the treatment groups ($\rho_0 = 0.39, \rho_1 = -0.19$).