Supplementary Materials for "A Maximum Likelihood Approach to Electronic Health Record Phenotyping using Positive and Unlabeled Patients"

Lingjiao Zhang, Xiruo Ding, Yanyuan Ma, Naveen Muthu, Imran Ajmal, Jason H. Moore, Daniel S. Herman, Jinbo Chen

Appendix A: Proof of Parameter Identifiability

The joint pdf of $(\mathbf{X}, S)$ can be derived as

$$f(\mathbf{X})\left\{\frac{hP(\mathbf{X};\boldsymbol{\beta})}{\int P(\mathbf{X};\boldsymbol{\beta})f(\mathbf{X})d\mathbf{X}}\right\}^{I(S=1)}\left\{1-\frac{hP(\mathbf{X};\boldsymbol{\beta})}{\int P(\mathbf{X};\boldsymbol{\beta})f(\mathbf{X})d\mathbf{X}}\right\}^{I(S=0)}$$

with unknown parameters $h, \boldsymbol{\beta}, f(\boldsymbol{X})$. Note that we have a random sample of $(\mathbf{X}, S)$'s, hence the observations $\mathbf{X}_i$'s form a random sample of $f(\mathbf{X})$. Thus, obviously, $f(\mathbf{X})$ is identifiable. Likewise, the observations S_i's form a random sample of the pmf of S, hence h is obviously identifiable. Assume that $\boldsymbol{\beta}$ is not identifiable. Then there exists $\boldsymbol{\beta}^*$ so that $p(\mathbf{X}, S; \boldsymbol{\beta}) = p(\mathbf{X}, S; \boldsymbol{\beta}^*)$ for all $(\mathbf{X}, S)$ pairs, which can be formalized as:

$$\begin{aligned}&\left\{\frac{hP(\boldsymbol{X};\boldsymbol{\beta})}{\int P(\boldsymbol{X};\boldsymbol{\beta})f(\boldsymbol{X})d\boldsymbol{X}}\right\}^{I(S=1)}\left\{1-\frac{hP(\boldsymbol{X};\boldsymbol{\beta})}{\int P(\boldsymbol{X};\boldsymbol{\beta})f(\boldsymbol{X})d\boldsymbol{X}}\right\}^{I(S=0)}\\ =\ &\left\{\frac{hP(\boldsymbol{X};\boldsymbol{\beta}^*)}{\int P(\boldsymbol{X};\boldsymbol{\beta}^*)f(\boldsymbol{X})d\boldsymbol{X}}\right\}^{I(S=1)}\left\{1-\frac{hP(\boldsymbol{X};\boldsymbol{\beta}^*)}{\int P(\boldsymbol{X};\boldsymbol{\beta}^*)f(\boldsymbol{X})d\boldsymbol{X}}\right\}^{I(S=0)}.\end{aligned}$$

Letting $S = 1$, this equation reduces to

$$\frac{hP(\boldsymbol{X};\boldsymbol{\beta})}{\int P(\boldsymbol{X};\boldsymbol{\beta})f(\boldsymbol{X})d\boldsymbol{X}} = \frac{hP(\boldsymbol{X};\boldsymbol{\beta}^*)}{\int P(\boldsymbol{X};\boldsymbol{\beta}^*)f(\boldsymbol{X})d\boldsymbol{X}},$$

or equivalently

$$\frac{P(\boldsymbol{X};\boldsymbol{\beta}^*)}{P(\boldsymbol{X};\boldsymbol{\beta})} = \frac{\int P(\boldsymbol{X};\boldsymbol{\beta}^*)f(\boldsymbol{X})d\boldsymbol{X}}{\int P(\boldsymbol{X};\boldsymbol{\beta})f(\boldsymbol{X})d\boldsymbol{X}}. \tag{1}$$

Note that the right hand side of (1) is a positive constant, hence the left hand side is also a constant. When the covariates $X_j \to \infty$ if $\beta_j > 0$ and $X_j \to -\infty$ if $\beta_j \leq 0$, regardless what the intercept term is, $P(\boldsymbol{X};\boldsymbol{\beta}) \to 1$ and $P(\boldsymbol{X};\boldsymbol{\beta}^*)$ converges to either 1 or 0. Thus, the left hand side goes to either 1 or 0. Because the right hand side is positive, hence we further conclude that the left hand side goes to 1. Because it has to be a constant, this implies that the left hand side is always 1, i.e. $P(\boldsymbol{X};\boldsymbol{\beta}^*) = P(\boldsymbol{X};\boldsymbol{\beta})$ at all $\boldsymbol{X}$. Thus $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, i.e. $\boldsymbol{\beta}$ is identifiable. Once $\boldsymbol{\beta}, h, f(\mathbf{X})$ are indentifiable, the disease prevalence $q = \int P(\boldsymbol{X};\boldsymbol{\beta})f(\boldsymbol{X})d\boldsymbol{X}$ is identifiable by definition.

Appendix B: Phenotyping with an anchor of varying sensitivity

Suppose that the population can be divided into K strata of size N_k, where the k^{th} stratum is indicated by $z_k,\ k = 1, ..., K$. Within each substratum, the anchor has a constant sensitivity. For notational convenience, Z is included as a component of predictors $\boldsymbol{X}$. The problem is then formalized as stratified conditional independence,

$$p(S = 1|Y = 1, \boldsymbol{X}) = p(S = 1|Y = 1, Z = z_k) = c_k, k = 1, ..., K.$$

The logistic model (2) remains the working model for prediction, where Z, if not predictive, has a log odds ratio equal to zero. The probability of observed data of an anchor-positive case and an unlabeled patient in each stratum is generalized as below to reflect variation in anchor sensitivity,

$$p(\boldsymbol{X}, S=1) = c_k p(Y=1|\boldsymbol{X}) p(\boldsymbol{X}),$$
$$p(\boldsymbol{X}, S=0) = \{1 - c_k p(Y=1|\boldsymbol{X})\} p(\boldsymbol{X}).$$

The likelihood function becomes the probability of observed data across all K strata,

$$l(\boldsymbol{\beta}, \{c_k\}) \propto \sum_{i=1}^{N} \sum_{k=1}^{K} I(Z_i = z_k)[S_i \log\{c_k P(\boldsymbol{X}_i; \boldsymbol{\beta})\} + (1 - S_i) \log\{1 - c_k P(\boldsymbol{X}_i; \boldsymbol{\beta})\}]$$

Parameter identifiability can be shown similarly as in the situation of constant anchor sensitivity. The MLE of $\boldsymbol{\beta}$ and $\{c_k,\ k = 1, ..., K\}$ can be obtained by maximizing the log likelihood function $l(\boldsymbol{\beta}, \{c_k\})$. Let $q_k \equiv p(Y = 1|z_k)$ denote the phenotype prevalence in the k^{th} stratum. Then q_k can be estimated as $\hat{h}_k/\hat{c}_k$, where $\hat{h}_k$ is the estimated anchor prevalence in the k^{th} stratum, $\hat{p}(S = 1|Z = z_k) = \sum_{i=1}^{N} I(Z_i = z_k, S_i = 1)/N_k$. The overall phenotype prevalence q can then be estimated as a weighted summation of q_k's, $\hat{q} = \sum_{k=1}^{K} N_k \hat{q}_k / N$.

Appendix C: Simulation Results

As shown in Table S1, the proposed maximum likelihood approach achieved consistent and efficient estimate of the odds ratio parameters $\boldsymbol{\beta}$, the anchor sensitivity c, and the phenotype prevalence q at different phenotype prevalence set-ups (5%, 10%, 15%, 20%).

To investigate the variation in model performance with respect to anchor sensitivity c, we considered two values for c, 0.5 and 0.2. With 10,000 observations in the training set and phenotype prevalence 10%, taking 0.5 as the decision threshold, the anchor variable helped identify 536 cases when c equaled 0.5 and 797 cases when c equaled 0.2. The estimates of c and q appeared to be consistent in both cases, although those at $c = 0.2$ had wider confidence intervals as expected (Table S2). The OR estimates $\hat{\boldsymbol{\beta}}$ appeared somewhat biased when $c = 0.2$. However, the bias largely disappeared when the size of the training set was increased to $20,000$ (Table S2).

To assess the model performance when the anchor sensitivity varies, we simulated a new population following the same steps as above, except that it consisted of two strata indicated by a binary variable Z that takes value 0 or 1 with probability $p(Z = 1) = 0.5$. Conditional on the phenotype the anchor variable was independent of all predictors within each strata, with $c_1 = p(S = 1|Y = 1, Z_1) = 0.2$, and $c_2 = p(S = 1|Y = 1, Z_2) = 0.8$. Applying the proposed method that takes the stratified conditional independence into account, the parameter estimates, $\hat{\boldsymbol{\beta}}$, $\hat{c_1}$, $\hat{c_2}$, $\hat{q}$, appeared consistent (Table S3). The anchor sensitivity c_1 was estimated to be 0.201 (SE: 0.025; 95% CI: 0.151, 0.251), c_2 was estimated to be 0.800 (SE: 0.022; 95% CI: 0.757, 0.843), and consequently the phenotype prevalence q was estimated to be 0.100 (SE: 0.004; 95% CI: 0.091, 0.109). On the other hand, if the variation in anchor sensitivity failed to be recognized, the fitted model can lead to biased parameter estimates (Table S3). To conclude, it is crucial to recognize the stratification, especially when the anchor sensitivities varies a lot.

Appendix D: Supplementary Tables

Table S1. Estimated odds ratio parameters $\boldsymbol{\beta}$, anchor sensitivity c, and phenotype prevalence q when the phenotype prevalence was equal to 5%, 10%, 15%, or 20%.

Phenotype prevalence		β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	c	q
5%													
	Mean	-2.629	0.208	0.433	0.634	-1.053	-1.472	1.899	-2.112	2.545	2.952	0.500	0.051
	Bias	-0.129	0.008	0.033	0.034	-0.053	-0.072	0.099	-0.112	0.145	0.152	0.0009	-0.0002
	SE	0.561	0.061	0.353	0.122	0.135	0.391	0.240	0.253	0.463	0.357	0.031	0.003
	ESE	0.573	0.062	0.368	0.127	0.144	0.416	0.258	0.269	0.476	0.383	0.031	0.003
10%													
	Mean	1.039	0.207	0.410	0.621	-1.031	-1.444	1.858	-2.067	2.482	2.895	0.500	0.100
	Bias	0.039	0.007	0.010	0.021	-0.031	-0.044	0.058	-0.067	0.082	0.095	0.0004	-0.00008
	SE	0.471	0.046	0.267	0.092	0.101	0.294	0.181	0.189	0.345	0.268	0.021	0.004
	ESE	0.486	0.047	0.271	0.091	0.106	0.301	0.192	0.197	0.350	0.284	0.021	0.004
15%													
	Mean	3.401	0.202	0.406	0.616	-1.023	-1.441	1.842	-2.049	2.461	2.868	0.500	0.148
	Bias	0.101	0.002	0.006	0.016	-0.023	-0.041	0.042	-0.049	0.061	0.068	0.0004	-0.0008
	SE	0.517	0.040	0.232	0.079	0.088	0.255	0.157	0.163	0.299	0.232	0.017	0.005
	ESE	0.520	0.041	0.237	0.082	0.091	0.258	0.166	0.171	0.308	0.237	0.017	0.005
20%													
	Mean	5.511	0.203	0.409	0.609	-1.019	-1.435	1.834	-2.039	2.443	2.856	0.500	0.204
	Bias	0.111	0.003	0.009	0.009	-0.019	-0.035	0.034	-0.039	0.043	0.056	0.0002	-0.0002
	SE	0.583	0.036	0.209	0.072	0.079	0.231	0.143	0.148	0.270	0.211	0.014	0.005
	ESE	0.562	0.036	0.214	0.072	0.078	0.227	0.141	0.144	0.268	0.210	0.014	0.005

Mean: the mean $\hat{\beta}$ estimate; Bias: the difference between the mean $\hat{\beta}$ and the true value of β; SE: the mean asymptotic standard error estimate; ESE: the empirical standard error estimate

Table S2. Estimated odds ratio parameters $\boldsymbol{\beta}$, anchor sensitivity c, and phenotype prevalence q at 10% phenotype prevalence when $c = 0.2$.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	c	q
Size of training set=10,000												
Mean	1.192	0.220	0.453	0.664	-1.115	-1.573	2.029	-2.243	2.698	3.143	0.201	0.100
Bias	0.192	0.020	0.053	0.064	-0.115	-0.173	0.229	-0.243	0.298	0.343	0.0007	0.0003
SE	0.889	0.085	0.484	0.173	0.202	0.544	0.363	0.383	0.658	0.540	0.017	0.006
ESE	1.422	0.109	0.554	0.189	0.296	0.831	0.737	0.687	1.292	1.085	0.017	0.007
Size of training set=20,000												
Mean	1.064	0.209	0.436	0.628	-1.044	-1.464	1.877	-2.087	2.495	2.925	0.200	0.100
Bias	0.064	0.009	0.036	0.028	-0.044	-0.064	0.077	-0.087	0.095	0.125	0.0005	-0.0003
SE	0.577	0.055	0.320	0.111	0.123	0.354	0.220	0.230	0.416	0.327	0.012	0.004
ESE	0.594	0.056	0.334	0.117	0.125	0.351	0.226	0.237	0.403	0.332	0.012	0.004

Mean: the mean $\hat{\beta}$ over 1000 iterations; Bias: the difference between the mean $\hat{\beta}$ and the true value of β; SE: the mean asymptotic standard error estimate; ESE: the empirical standard error.

Table S3. Estimated ($\hat{\beta}$, $\hat{c_1}$, $\hat{c_2}$, $\hat{q}$) for stratified conditional independence with $c_1 = 0.2$ and $c_2 = 0.8$ at 10% phenotype prevalence.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	c_1	c_2	q
Varying anchor sensitivity[1]													
Mean	1.046	0.207	0.410	0.622	-1.031	-1.437	1.856	-2.062	2.478	2.882	0.201	0.800	0.100
Bias	0.046	0.007	0.010	0.022	-0.031	-0.037	0.056	-0.062	0.078	0.082	0.001	0.000	-0.000
ASE	0.559	0.041	0.238	0.081	0.089	0.262	0.159	0.165	0.459	0.234	0.025	0.022	0.004
ESE	0.577	0.040	0.243	0.082	0.091	0.258	0.161	0.168	0.476	0.237	0.025	0.022	0.004
Constant anchor sensitivity[2]													
Mean	-2.338	0.148	0.293	0.445	-0.735	-1.025	1.324	-1.471	4.827	2.047	0.712	_	0.080
Bias	-3.338	-0.052	-0.107	-0.155	0.265	0.375	-0.476	0.529	2.427	-0.753	0.212	_	-0.021
ASE	0.363	0.033	0.195	0.064	0.065	0.210	0.116	0.118	0.327	0.168	0.028	_	0.004
ESE	0.504	0.034	0.200	0.071	0.077	0.212	0.138	0.145	0.349	0.209	0.036	_	0.005

Mean: the mean estimates over 1000 iterations; Bias: the difference between the mean estimates and the true values; SE: the mean asymptotic standard error estimate; ESE: the empirical standard error. [1]: the fitted model recognized the variation in anchor sensitivity; [2]: the fitted model failed to recognize the variation in anchor sensitivity.

Table S4. Variables of interest for primary care patients identifying Penn Medicine hypertension patients with hypokalemia requiring oral supplementation

Variable	$N = 10000$
Mean of systolic blood pressure measured[1]	131.5 (9.5)
Minimum of potassium lab test[1]	3.9 (0.4)
Maximum of potassium lab test[1]	4.7 (0.5)
1st quantile of potassium lab test[1]	4.1 (0.3)
3rd quantile of potassium lab test[1]	27.7 (2.2)
3rd quantile of potassium lab test[1]	4.4 (0.4)
Median of potassium lab test[1]	4.3 (0.3)
Number of low potassium test results[2]	0 (0-44)
Number of Hypokalemia Diagnosis[2]	0 (0-78)
Sum of E87.6 ICD-9 and ICD-10 codes[2]	0 (0-98)
Sum of I10 ICD-9 and ICD-10 codes[2]	19 (0-296)
Sum of I16.1 ICD-9 and ICD-10 codes[2]	2 (0-244)

[1]: Mean (SD), [2]: Median (Minimum-Maximum)

Table S5. Characteristics of patients in Penn Medicine EHR that had an order for a PA screening laboratory test

	Final population ($N = 6319$)
Age[1]	54 (16)
Gender	
Male (%)	2471 (39)
Female (%)	3848 (61)
Race	
Caucasian (%)	3270 (52)
African-American (%)	2440 (39)
Other (%)	609 (9)
Aldosterone[2] (ng/DL)	8.2 (4.4-14.7)
Renin[2] (ng/mL/hr)	0.9 (0.3-2.7)
$Aldo : Renin$[2]	8.0 (2.6-27.8)
Potassium[1] (mmol/L)	4.1 (0.6)
Sodium[1] (mmol/L)	139 (3.1)
$CO2$[1] (mmol/L)	26.6 (3.4)

[1]: Mean (SD), [2]: Median (IQR)

Table S6. Dictionary for variables of interest for identifying Penn Medicine hypertension patients with primary aldosteronism

	Variable	Description
Demographics	age	Age when aldosterone or renin test was performed (year)
	gender	Gender
	race	Race (Black/White/Other)
	hisp	Hispanic (Yes/No)
Pre-visit	dbp	Diastolic blood pressure, from office visit closest (<= 14 days) to aldosterone/renin testing
	sbp	Systolic blood pressure, from office visit closest (<= 14 days) to aldosterone/renin testing
	time_bp_to_1st_RAR_yr	Time interval (years) between first office visit with blood pressure recorded to aldosterone/renin test
	time_enc_to_1st_RAR_yr	Time interval (years) between first clinical encounter to aldosterone/renin test
Laboratory data	aldo	Serum aldosterone concentration (ng/dL)
	pra	Plasma renin activity (ng/mL/hr)
	aldo:pra	The aldosterone:renin ratio ((ng Aldosterone/dL)/(ng Angiotensin II/mL/hr))
	test_potassium	Blood potassium concentration (mmol/L)
	test_sodium	Blood sodium concentration (mmol/L)
	test_carbon_dioxide	Blood carbon dioxide concentration (mmol/L)
Encounter	enc_n	Number of clinical encounters
	enc_bp_n	Number of office visits with blood pressure recorded
	time_bp_after_1st_RAR_yr	Time interval (years) between aldosterone/renin test and last office visit with blood pressure
	time_enc_after_1st_RAR_yr	Time interval (years) between aldosterone/renin test and last clinical encounter
		Sum of the number of encounters with primary aldosteronism diagnosis codes
Diagnoses	Dx_h2_E26.0_9_n	(255.1, 255.10, 255.11, 255.12, E26.0, E26.01, E26.02, E26.09, E26.9)
		Sum of the number of encounters with other hyperaldosteronism diagnosis codes
	Dx_h2_E26.1_8_n	(255.13, 255.14, E26.1, E26.81, E26.89)
Notes	re_hyperaldo	count of 'hyperaldo' mentions in clinical notes
	re_primaryaldo	count of 'primary aldo' mentioned in the clinical notes
	re_bah	count of 'bah' mentioned in the clinical notes
	re_adrenal_adenoma	count of 'adrenal_adenoma' mentioned in the clinical notes
	re_htn	count of 'hypertension' mentioned in the clinical notes
	re_adrenalectomy	count of 'adrenalectomy' mentioned in the clinical notes

Table S7. Final phenotyping models for identifying Penn Medicine hypertension patients with primary aldosteronism

	Case set A			Case set B		
	Coefficient	SE	p_value	Coefficient	SE	p_value
Intercept.	-9.06	2.20	0.00	-9.35	2.17	0.00
age	-0.43	0.23	0.06	-0.38	0.20	0.06
gender	0.39	0.37	0.29	0.40	0.33	0.23
race	0.59	0.22	0.01	0.56	0.21	0.01
hisp	-0.12	1.16	0.91	-1.07	1.16	0.35
dbp	-0.31	0.25	0.22	-0.42	0.23	0.07
sbp	0.32	0.27	0.24	0.41	0.24	0.09
time_bp_to_1st_RAR_yr	-0.32	0.31	0.29	-0.16	0.27	0.54
time_enc_to_1st_RAR_yr	-0.20	0.39	0.60	-0.23	0.32	0.48
aldo	0.80	0.39	0.04	0.94	0.38	0.01
pra	0.03	0.61	0.96	-0.05	0.61	0.94
aldo:pra	1.13	0.66	0.09	1.00	0.66	0.13
potassium_indicator	2.89	1.37	0.03	2.96	1.30	0.02
potassium_test	0.43	0.69	0.53	-0.10	0.64	0.88
sodium_indicator	9.08	1.37	0.00	8.85	1.33	0.00
sodium_test	2.82	3.56	0.43	3.78	3.52	0.28
carbon_dioxide_indicator	-10.46	1.37	0.00	-9.82	1.33	0.00
carbon_dioxide_test	-0.33	0.58	0.57	-0.27	0.59	0.64
enc_n	0.80	0.48	0.10	0.25	0.46	0.59
enc_bp_n	-1.53	0.44	0.00	-1.16	0.42	0.01
time_enc_after_1st_RAR_yr	0.53	0.31	0.08	0.28	0.28	0.33
time_bp_after_1st_RAR_yr	-0.12	0.33	0.71	-0.04	0.33	0.90
Dx_h2_E26.0_9_n	0.53	0.16	0.00	1.22	0.23	0.00
Dx_h2_E26.1_8_n	0.17	0.07	0.01	0.17	0.06	0.01
re_hyperaldo	0.38	0.20	0.05	0.39	0.25	0.12
re_primary_aldo	0.13	0.08	0.12	0.54	0.18	0.00
re_bah	1.53	0.53	0.00	2.10	0.49	0.00
re_adrenal_adenoma	0.52	0.16	0.00	0.49	0.14	0.00
re_htn	-0.21	0.32	0.50	-0.24	0.34	0.47
re_adrenalectomy	0.82	0.18	0.00	0.90	0.16	0.00

Table S8. Estimated anchor sensitivity and prevalence, Mean (95% CI), using stratified vs. non-stratified ML method for identifying PA patients

	Stratified			Non-Stratified	
	c^1	c^2	q	c	q
Set A					
enc_n	0.69 (0.54, 0.84)	0.49 (0.38, 0.6)	0.040 (0.033, 0.048)	0.56 (0.46, 0.65)	0.042 (0.034, 0.0
enc_bp_n	0.72 (0.52, 0.92)	0.51 (0.4, 0.62)	0.039 (0.031, 0.047)	-	-
time_enc_after_1st_RAR_yr	0.46 (0.31, 0.6)	0.59 (0.49, 0.69)	0.044 (0.035, 0.053)	-	-
time_bp_after_1st_RAR_yr	0.49 (0.33, 0.64)	0.59 (0.47, 0.7)	0.043 (0.034, 0.052)	-	-
Set B					
enc_n	0.76 (0.65, 0.86)	0.52 (0.43, 0.61)	0.049 (0.042, 0.056)	0.62 (0.54, 0.69)	0.051 (0.044, 0.0
enc_bp_n	0.72 (0.61, 0.84)	0.56 (0.46, 0.65)	0.050 (0.043, 0.057)	-	-
time_enc_after_1st_RAR_yr	0.75 (0.64, 0.87)	0.54 (0.45, 0.63)	0.051 (0.043, 0.058)	-	-
time_bp_after_1st_RAR_yr	0.76 (0.64, 0.87)	0.53 (0.44, 0.62)	0.051 (0.044, 0.058)	-	-

[1]Estimated anchor sensitivity in the 1st strata, i.e., variable value < median

[2]Estimated anchor sensitivity in the 2nd strata, i.e., variable value >= median

Table S9. Estimated regression coefficients β using stratified vs. non-stratified ML method for identifying PA patients

	Set A					Set B				
	Non-S	S_1	S_2	S_3	S_4	Non-S	S_1	S_2	S_3	S_4
(Intercept)	-9.06	-8.59	-8.58	-9.31	-9.11	-9.35	-8.95	-9.17	-9.18	-9.41
age	-0.43	-0.45	-0.44	-0.45	-0.43	-0.38	-0.39	-0.38	-0.36	-0.41
gender	0.39	0.36	0.37	0.39	0.4	0.4	0.4	0.4	0.44	0.42
race	0.59	0.57	0.55	0.6	0.59	0.56	0.56	0.56	0.57	0.58
hisp	-0.12	-0.16	-0.11	-0.12	-0.14	-1.07	-1.31	-1.23	-1.23	-1.27
dbp	-0.31	-0.31	-0.31	-0.31	-0.3	-0.42	-0.42	-0.42	-0.37	-0.39
sbp	0.32	0.3	0.3	0.33	0.32	0.41	0.37	0.41	0.36	0.38
time_bp_to_1st_RAR_yr	-0.32	-0.31	-0.31	-0.32	-0.34	-0.16	-0.13	-0.14	-0.13	-0.1
time_enc_to_1st_RAR_yr	-0.2	-0.19	-0.21	-0.21	-0.19	-0.23	-0.23	-0.23	-0.26	-0.29
aldo	0.8	0.72	0.72	0.85	0.82	0.94	0.81	0.86	0.86	0.88
pra	0.03	0.08	0.04	-0.02	0.01	-0.05	0.04	-0.01	0.02	0.02
aldo:pra	1.13	1.12	1.08	1.08	1.1	1.0	1.01	0.99	1.03	1.05
potassium_indicator	2.89	2.53	2.48	3.15	2.84	2.96	2.44	2.66	2.54	3.08
potassium_test	0.43	0.31	0.29	0.43	0.43	-0.1	-0.22	-0.18	-0.1	-0.12
sodium_indicator	9.08	9.1	9.15	9.1	9.17	8.85	8.99	8.95	8.95	8.76
sodium_test	2.82	2.35	2.34	3.32	2.99	3.78	3.36	3.6	3.43	3.74
carbon_dioxide_indicator	-10.46	-10.43	-10.39	-10.43	-10.36	-9.82	-9.69	-9.72	-9.73	-9.92
carbon_dioxide_test	-0.33	-0.36	-0.34	-0.33	-0.33	-0.27	-0.22	-0.23	-0.25	-0.26
enc_n	0.8	0.95	0.86	0.77	0.77	0.25	0.44	0.3	0.27	0.25
enc_bp_n	-1.53	-1.5	-1.38	-1.55	-1.49	-1.16	-1.15	-1.07	-1.13	-1.2
time_enc_after_1st_RAR_yr	0.53	0.56	0.54	0.35	0.53	0.28	0.31	0.3	0.52	0.26
time_bp_after_1st_RAR_yr	-0.12	-0.19	-0.21	-0.05	-0.22	-0.04	-0.12	-0.09	-0.11	0.22
Dx_h2_E26.0_9_n	0.53	0.53	0.53	0.53	0.52	1.22	1.21	1.23	1.28	1.29
Dx_h2_E26.1_8_n	0.17	0.16	0.16	0.17	0.17	0.17	0.15	0.16	0.16	0.16
re_hyperaldo	0.38	0.39	0.38	0.39	0.38	0.39	0.43	0.42	0.4	0.42
re_primary_aldo	0.13	0.12	0.11	0.14	0.14	0.54	0.54	0.55	0.5	0.49
re_bah	1.53	1.47	1.38	1.49	1.5	2.1	1.93	1.95	2.02	2
re_adrenal_adenoma	0.52	0.53	0.53	0.53	0.51	0.49	0.48	0.49	0.48	0.5
re_htn	-0.21	-0.18	-0.19	-0.22	-0.22	-0.24	-0.18	-0.22	-0.24	-0.24
re_adrenalectomy	0.82	0.79	0.75	0.85	0.83	0.9	0.89	0.88	0.87	0.88

Non-S: non-stratified

S_1-S_4: stratified based on enc_n, enc_bp_n, $time\_enc\_after\_1st\_RAR\_yr$, $time\_bp\_after\_1st\_RAR\_yr$ respectively