HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

The CAUSALab

Effects of intergenerational exposure interventions on adolescent outcomes

An application of IPW to prebirth cohort data

Yu-Han Chiu, MD, ScD

Department of Epidemiology

Harvard T.H. Chan School of Public Health

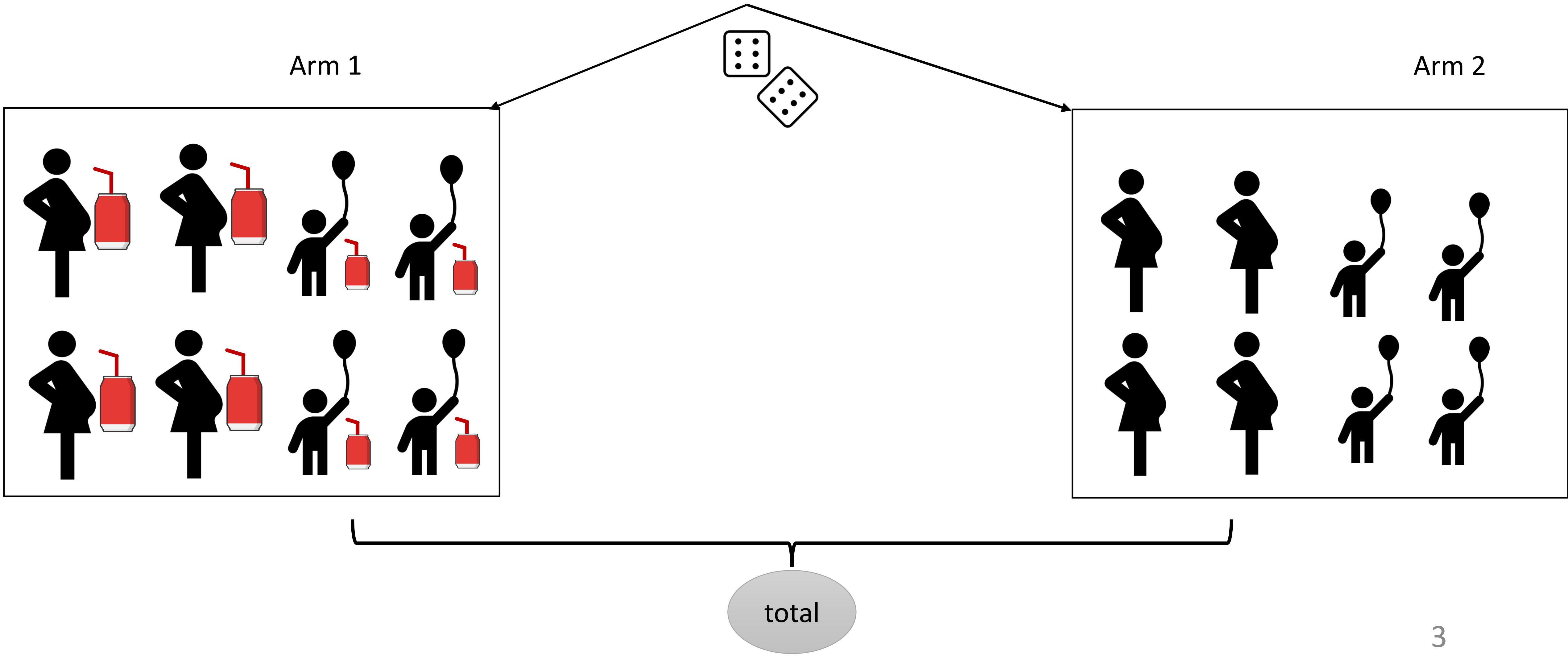
yuc187@mail.Harvard.edu

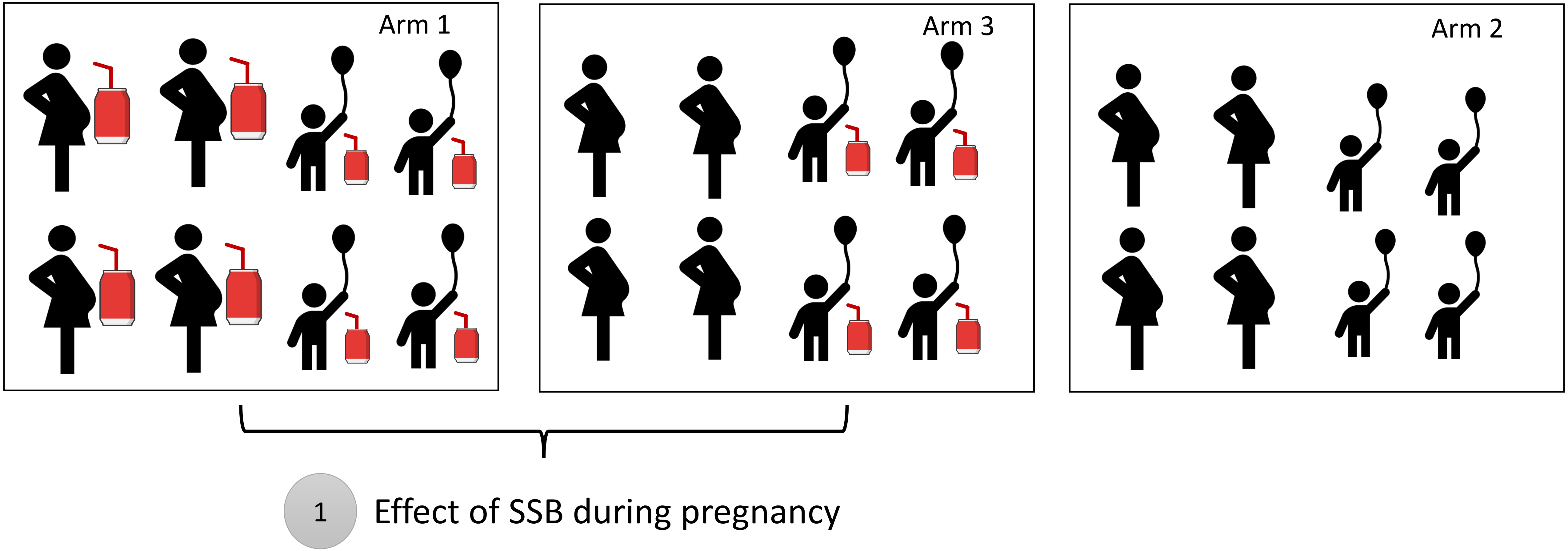
@HanChiu_

Sustained interventions on SSB consumption and adolescent BMI

- Sugar-sweetened beverage (SSB): a prime target of obesity prevention
- American Academy of Pediatrics:
  - Limiting SSB intake to 6 oz per day (0.5 cups/day) for 1 to 6 years
- **Study question:** Will maintaining SSB intake within the recommended ranges throughout prenatal and postnatal periods reduce BMI Z-score in early adolescence ?
  - If so, if there is a sensitive window

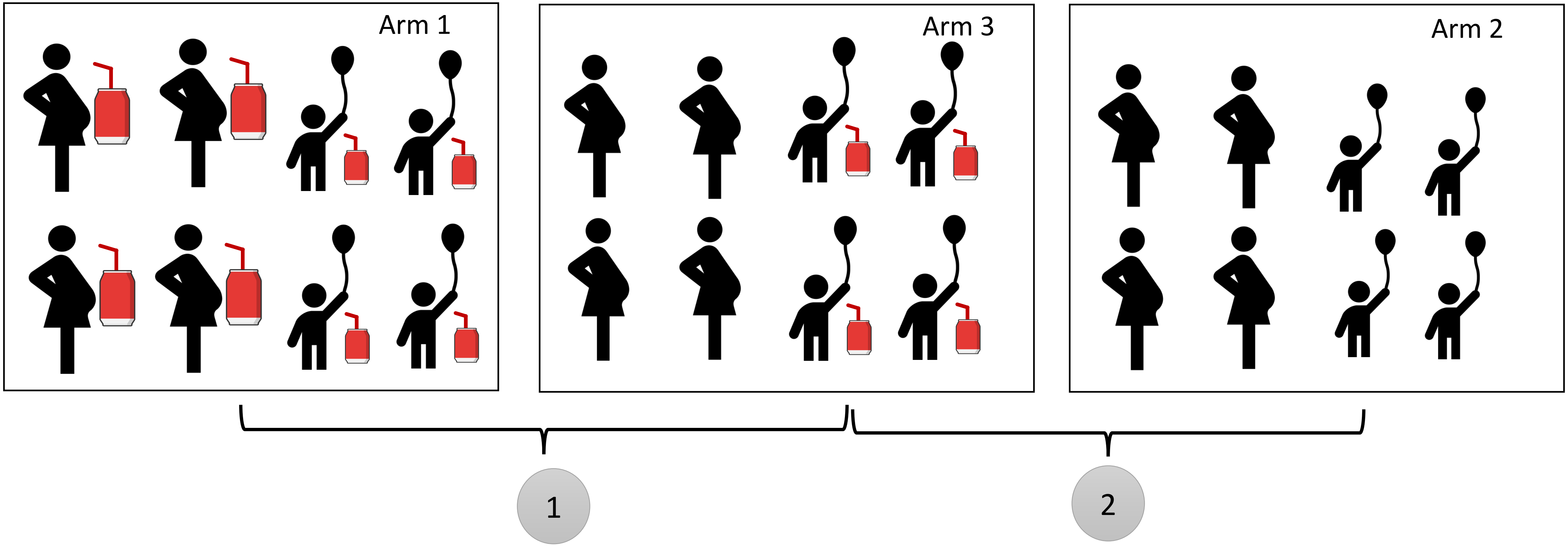
How to test this in a randomized trial?





Effect of SSB during pregnancy

*Assuming no interaction



total = 1 + 2

If 1 ≫ 2 Sensitive window during pregnancy

*Assuming no interaction

Protocol of the target trial

Eligibility Criteria	Pregnant women at first OB visit and enrolled before 24 weeks of gestation
Dietary Strategies	1. Maintain SSB intake below the recommended range (0.5 cup/day) 2. Maintain SSB intake above the recommended range (0.5 cup/day) through pregnancy (early, late) and childhood (3,4,5,6 years) 3. Maintain SSB intake below the recommended range in pregnancy and above the recommended range in childhood
Assignment	Participants are randomly assigned to one of the three strategies
Outcome	BMI Z score in early teen visit (12 years)
Follow up	Start at assignment until loss to follow-up, drop out, fetal loss, or early teen visit, whichever comes first
Causal contrast	Per protocol effect (i.e., the effect if everyone had adhered to the assigned intervention at all times)

Protocol of the target trial

Eligibility Criteria	Pregnant women at first OB visit and enrolled before 24 weeks of gestation
Dietary Strategies	1. Maintain SSB intake below the recommended range (0.5 cup/day) 2. Maintain SSB intake above the recommended range (0.5 cup/day) through pregnancy (early, late) and childhood (3,4,5,6 years) 3. Maintain SSB intake below the recommended range in pregnancy and above the recommended range in childhood
Assignment	Participants are randomly assigned to one of the three strategies
Outcome	BMI Z score in early teen visit (12 years)
Follow up	Start at assignment until loss to follow-up, drop out, fetal loss, or early teen visit, whichever comes first
Causal contrast	Per protocol effect (i.e., the effect if everyone had adhered to the assigned intervention at all times)

Protocol of the target trial

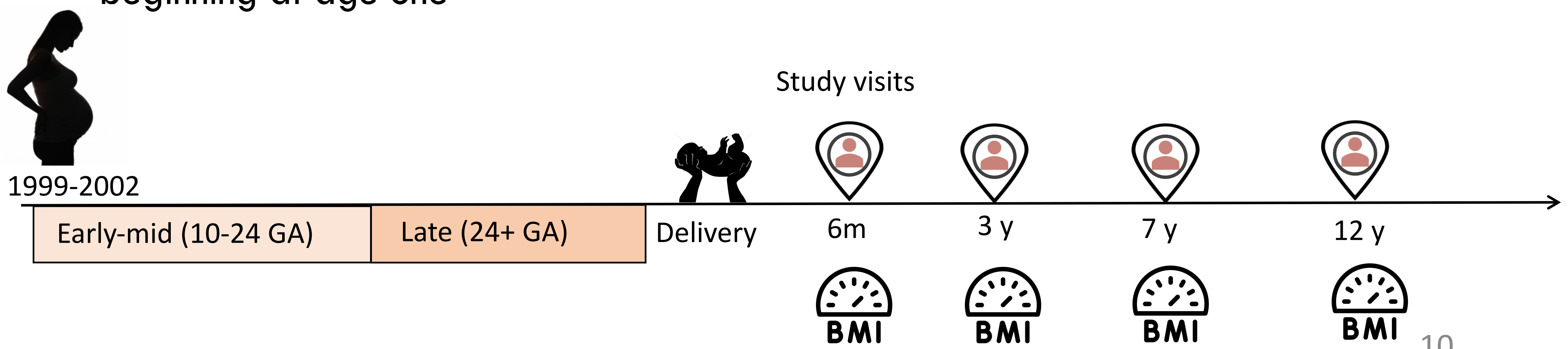
Eligibility Criteria	Pregnant women at first OB visit and enrolled before 24 weeks of gestation
Dietary Strategies	1. Maintain SSB intake below the recommended range (0.5 cup/day) 2. Maintain SSB intake above the recommended range (0.5 cup/day) through pregnancy (early, late) and childhood (3,4,5,6 years) 3. Maintain SSB intake below the recommended range in pregnancy and above the recommended range in childhood
Assignment	Participants are randomly assigned to one of the three strategies
Outcome	BMI Z score in early teen visit (12 years)
Follow up	Start at assignment until loss to follow-up, drop out, fetal loss, or early teen visit, whichever comes first
Causal contrast	Per protocol effect (i.e., the effect if everyone had adhered to the assigned intervention at all times)

Protocol of the target trial

Eligibility Criteria	Pregnant women at first OB visit and enrolled before 24 weeks of gestation
Dietary Strategies	1. Maintain SSB intake below the recommended range (0.5 cup/day) 2. Maintain SSB intake above the recommended range (0.5 cup/day) through pregnancy (early, late) and childhood (3,4,5,6 years) 3. Maintain SSB intake below the recommended range in pregnancy and above the recommended range in childhood
Assignment	Participants are randomly assigned to one of the three strategies
Outcome	BMI Z score in early teen visit (12 years)
Follow up	Start at assignment (baseline visit) until loss to follow-up, drop out, fetal loss, or early teen visit, whichever comes first
Causal contrast	Per protocol effect (i.e., the effect if everyone had adhered to the assigned intervention at all times)

Observational data: Project Viva

- A prospective pre-birth cohort
- Enrolled participants during initial obstetric care visit (~10 GA) from Harvard Vanguard Medical Associates in Eastern Massachusetts
- **In-person study visits** at 6 mon, 3 y, 7 y, 12 y
- **Diet Assessment:** early-mid pregnancy, late pregnancy, and annual questionnaire beginning at age one



	The protocol of target trial	Target trial emulation
Eligibility Criteria	Pregnant women at first OB visit, enroll before before 24 weeks of gestation	Same plus one requirement: completing a baseline questionnaire and a dietary questionnaire
Dietary Strategies	1. Maintain SSB below 0.5 servings/day 2. Maintain SSB above 0.5 servings/day through pregnancy and childhood 3-6 years 3. Maintain SSB below 0.5 in pregnancy and above 0.5 servings in childhood	Same
Assignment	Participants are randomly assigned to one of the three strategies	Emulate randomization by adjusting baseline confounders
Outcome	BMI Z score in early adolescence	Same
Follow up	Start at assignment (baseline) until loss to follow-up, drop out, fetal loss, or early adolescence visit, whichever comes first	Same. • Baseline is defined as completion of a baseline questionnaire • Loss to follow up: not returning questionnaires or study visits

Hands-on analysis

Notations

- Let t denote a measurement interval
  - t=0: the early pregnancy
  - t=1: the late pregnancy
  - t=2: 3 years
  - t=3: 4 years
  - t=4: 5 years
  - t=5: 6 years
- L_t is a vector of covariates measured in interval t
- V is a component of the vector of baseline covariates L_0
- A_t is SSB intake in interval t
- R_t is indicator of SSB intake $\leq$ 0.5 servings/day in interval t
  - R_t=1 if $A_t \leq 0.5$ and R_t=0 if $A_t > 0.5$
- C_{t+1} is an indicator of censoring by interval $t + 1$

Step 1. Data set construction

A wide format

id	R_0	R_1	R_2	R_3	R_4	R_5	Y	
1	1	1	0	0	0	1	1.02	
2	0	1	0	0	1	1	-0.07	
3	1	1	1	.	.	.	.	
.								
.								
1584								

A long format dataset allows for pooled over time models

$R_t \sim I(intervaal) + \alpha_1 * R_{t-1} + \alpha_2 * L_t + \cdots$

We chose a wide format here

- covariate distributions vary between moms and children at different age

$R_1 \sim \gamma_1 * R_0 + \gamma_2 * L_0 + \cdots$

$R_2 \sim \theta_1 * R_1 + \theta_2 * L_1 + \cdots$

A long format

id	t	R_t	Y
1	0	1	
1	1	1	
1	2	0	
1	3	0	
1	4	0	
1	5	1	1.02
2	0	0	
2	1	1	
2	2	0	
2	3	0	
2	4	1	
2	5	1	-0.07
3	1	1	
3	2	1	
3	3	1	

Step 1. Data set construction

- Assuming temporal order (C_t, L_t, A_t)
- Define C_{t+1} as an indicator of censoring by interval $t + 1$

id	R_0	R_1	R_2	R_3	R_4	R_5	C_1	C_2	C_3	C_4	C_5	C_6
51	1	1	0	0	0	1	0	0	0	0	0	0
52	1	1	1	.	.	.	0	0	1			
.												
.												

If $C_3 = 1$ then $\underline{L}_3$ or $\underline{A}_3$ are missing

Step 2a. Censoring weight

$$W^c = \frac{\boxed{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{R}_t, V)}}{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{A}_t, \bar{L}_t)}$$

Fit a logistic regression model to predict probability of $\boldsymbol{C_{t+1} = 0}$

- Restrict to those who **remained uncensored through $\boldsymbol{t}$**
- ~her/his SSB range history ($\boldsymbol{\bar{R}_t}$) and baseline covariates $\boldsymbol{V}$

Step 2a. Censoring weight

$$W^c = \frac{\boxed{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{R}_t, V)}}{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{A}_t, \bar{L}_t)}$$

Fit a logistic regression model to predict probability of $\boldsymbol{C_{t+1} = 0}$

- Restrict to those who **remained uncensored through $\boldsymbol{t}$**
- ~her/his SSB range history ($\bar{\boldsymbol{R}}_t$) and baseline covariates $\boldsymbol{V}$

```
proc logistic data=bytimesuse  ;
model cen1 =  &V  Rt_f1   ;
output out=cweightn1 p=pcn1;
run;

proc logistic data=bytimesuse   ;
where cen1=0;
model cen2 =  &V  Rt_f2  Rt_f1 ;
output out=cweightn2 p=pcn2;
run;

proc logistic data=bytimesuse  ;
where cen2=0;
model cen3 =  &V   Rt_3y  Rt_f2 ;
output out=cweightn3 p=pcn3;
run;
```

pcn3: estimated conditional probability of not censored by t=3

Step 2a. Censoring weight

$$W^c = \frac{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{R}_t, V)}{\boxed{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{A}_t, \bar{L}_t)}}$$

Fit a logistic regression model to predict probability of $\boldsymbol{C_{t+1} = 0}$

- Restrict to those who **remained uncensored through $\boldsymbol{t}$**
- ~a user-chosen function of SSB history ($\boldsymbol{\bar{A}_t}$) and confounder history ($\boldsymbol{\bar{L}_t}$)

Step 2a. Censoring weight

$$W^c = \frac{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{R}_t, V)}{\boxed{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{A}_t, \bar{L}_t)}}$$

Fit a logistic regression model to predict probability of $\boldsymbol{C_{t+1} = 0}$

- Restrict to those who **remained uncensored through** $\boldsymbol{t}$
- ~ SSB history ($\boldsymbol{\bar{A}_t}$) and confounder history ($\boldsymbol{\bar{L}_t}$)

A user-chosen function of past SSB ($\bar{A}_2$): Rt_3y + cumulative average of SSB during early& late pregnancy

```
proc logistic data=cweightn3  ;
model cen3 = Rt_3y    qssb_cumavgM1 qssb_cumavgM2  &V
qcalor_cumavgM1 qcalor_cumavgM2  qbwz1  qbwz2  qcalor_ch21 qcalor_ch22  TV_3ycat_1 ;
output out=cweight3 p=pcd3;
run;
```

History of baseline + post-baseline covariates ($\bar{L}_2$): caloric intake during pregnancy, birth weight, caloric intake at 2 years, TV at 3 years, and L_0 $(\&V)$

Step 2a. Censoring weight

For uncensored participant : $$W^c = \frac{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{R}_t, V)}{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{A}_t, \bar{L}_t)}$$

```
cwght_n = pcn1*pcn2*pcn3*pcn4*pcn5*pcn6;
cwght_d = pcd1*pcd2*pcd3*pcd4*pcd5*pcd6;
cwght= cwght_n / cwght_d ;
```

For censored participant : $$W^c = 0$$

Step 2b. Exposure weight

$$W_i^{\bar{r}} = \frac{\boxed{\prod_{t=0}^{t=5} P(R_t|\bar{C}_t = 0, \bar{R}_{t-1}, V) I(R_t = r_t)}}{\prod_{t=0}^{t=5} P(R_t|\bar{C}_t = 0, \bar{A}_{t-1}, \bar{L}_t)}$$

If her/his value is 1, this is probability that $R_t = 1$
If her/his value is 0, this is probability that $R_t = 0$

Fit a logistic regression model to predict $R_t = 1$

- Restrict to those **remained uncensored**
- As a function of past history of SSB range ($\bar{\boldsymbol{R}}_{\boldsymbol{t-1}}$) and baseline covariates $\boldsymbol{V}$

Step 2b Exposure weight

$$W_i^{\bar{r}} = \frac{\boxed{\prod_{t=0}^{t=5} P(R_t|\bar{C}_t = 0, \bar{R}_{t-1}, V) I(R_t = r_t)}}{\prod_{t=0}^{t=5} P(R_t|\bar{C}_t = 0, \bar{A}_{t-1}, \bar{L}_t)}$$

Fit a logistic regression model to predict $R_t = 1$

- Restrict to those **remained uncensored**
- As a function of past history of SSB range ($\bar{\boldsymbol{R}}_{\boldsymbol{t-1}}$) and baseline covariates $\boldsymbol{V}$

```
*Numerator of treatment/exposure weight;
```

```
proc logistic descending data=bytimes  noprint;
model Rt_f1=  &V             ;
freq numberhits;
output out=tweightn0 p=ptn0;
run;
```

```
proc logistic descending data=bytimes  noprint;
where cen1=0;
model Rt_f2=  &V    Rt_f1        ;
freq numberhits;
output out=tweightn1 p=ptn1;
run;
```

```
proc logistic descending data=bytimes  noprint;
where cen2=0;
model Rt_3y=   &V   Rt_f2       ;
freq numberhits;
output out=tweightn2 p=ptn2;
run;
```

Step 2b Exposure weight

$$W_i^{\bar{r}} = \frac{\prod_{t=0}^{t=5} P(R_t | \bar{C}_t = 0, \bar{R}_{t-1}, V) I(R_t = r_t)}{\boxed{\prod_{t=0}^{t=5} P(R_t | \bar{C}_t = 0, \bar{A}_{t-1}, \bar{L}_t)}}$$

Fit a logistic regression model to predict $R_t = 1$

- Restrict to those **remained uncensored**
- As a function of a user-chosen function of past SSB ($\bar{\boldsymbol{A}}_{\boldsymbol{t-1}}$), baseline covariates $\boldsymbol{V}$, and past history of time-varying covariates ($\bar{\boldsymbol{L}}_{\boldsymbol{t}}$)

$\bar{\boldsymbol{A}}_{\boldsymbol{1}}$: tertiles of average SSB intake during early and late pregnancy

```
proc logistic descending data=tweightn2 ;
model Rt_3y = qssb_cumavgM1 qssb_cumavgM2 &V
qcalor_cumavgM1 qcalor_cumavgM2  qbwz1  qbwz2  qcalor_ch21 qcalor_ch22  TV_3ycat_1  ;
output out=tweight2 p=ptd2;
run;
```

$\bar{\boldsymbol{L}}_{\boldsymbol{2}}$: birth weight + caloric intake at year 2 + TV watching at age 3 + &V

Step 2b Exposure weight

$$W_i^{\bar{r}} = \frac{\prod_{t=0}^{t=5} P(R_t|\bar{C}_t = 0, \bar{R}_{t-1}, V) I(R_t = r_t)}{\prod_{t=0}^{t=5} P(R_t|\bar{C}_t = 0, \bar{A}_{t-1}, \bar{L}_t)}$$

What we want: the probability that the individual received his/her own observed exposure at t

What models predict: the probability that R_t=1 for everyone

1

```
ptn_0=Rt_f1*ptn0+(1-Rt_f1)*(1-ptn0);
ptd_0=Rt_f1*ptd0+(1-Rt_f1)*(1-ptd0);
```

2

```
twght_n= ptn_0* ptn_1* ptn_2* ptn_3* ptn_4* ptn_5;
twght_d= ptd_0* ptd_1* ptd_2* ptd_3* ptd_4* ptd_5;
twght= twght_n/ twght_d;
```

Step 2 Overall weight

For uncensored participant:

$$W^{c} = \frac{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{R}_t, V)}{\prod_{t=0}^{t=5} P(C_{t+1} = 0 | \bar{C}_t = 0, \bar{A}_t, \bar{L}_t)}$$

$$W^{\bar{\gamma}} = \frac{\prod_{t=0}^{t=5} P(R_t | \bar{C}_t = 0, \bar{R}_{t-1}, V) I(R_t = r_t)}{\prod_{t=0}^{t=5} P(R_t | \bar{C}_t = 0, \bar{A}_{t-1}, \bar{L}_t)}$$

$$W = W^{c} X\, W^{\bar{\gamma}}$$

For censored participant:

$$W = 0$$

Step 2 Truncate extreme weight (to 99th percentile)

```
proc univariate data=weights0 noprint;
var stabw ;
freq numberhits;
output out = pctlweights
pctlpre = wgt_
pctlname = ulim99
pctlpts = 99
  ;

data weights;
set weights;
if stabw>wgt_ulim99 then stabw=wgt_ulim99;
run;
```

Step 3 Weighted outcome regression

$$E[Y^{\bar{r}_K,\bar{c}_{K+1}=0}|V] = b_{\circ}+\beta_0 R_{t=0} + \beta_1 R_{t=1}+\beta_2 R_{t=2}+\beta_3 R_{t=3}+\beta_4 R_{t=4} + \beta_5 R_{t=5}+\text{baseline V}$$

```
proc reg  data=bytimes5 outest=outc   ;
model BMIZ_ET = Rt_f1  Rt_f2   Rt_3y Rt_4y  Rt_5y Rt_6y  &V ;
weight stabw;
run;
```

Step 4: Estimating the intervention mean differences

- $E[Y^{\bar{r}_K, \bar{c}_{K+1}=0}|V] = b_o + \beta_0 R_{t=0} + \beta_1 R_{t=1} + \beta_2 R_{t=2} + \beta_3 R_{t=3} + \beta_4 R_{t=4} + \beta_5 R_{t=5} + V$
- IPW estimates of $\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$ interpreted as difference in mean BMI-Z
  - ❑had everyone consumed SSB below vs above the thresholds at all times t
  - ❑had everyone consumed juice above the thresholds at all times t
- 95%CI via bootstraps

Note: In our example, we consider a linear function (e.g., BMI Z score) and no effect modification by V, we can simply take the sum of the coefficients from each Rt
For general case, see slide 30-35

Step 4: Using betas to assess the sensitivity windows

- $\mathrm{E}[Y^{\bar{r}_K, \bar{c}_{K+1}=0} | V] = b_o + \beta_0 R_{t=0} + \beta_1 R_{t=1} + \beta_2 R_{t=2} + \beta_3 R_{t=3} + \beta_4 R_{t=4} + \beta_5 R_{t=5} + V$
- $\beta_0 + \beta_1$: difference in mean BMI-Z had everyone consumed SSB below (vs above) the thresholds at pregnancy (t=0,1)
- $\beta_2 + \beta_3 + \beta_4 + \beta_5$: difference in mean BMI-Z had everyone consumed SSB below (versus above) the thresholds during early childhood (t=3, 4, 5, 6)

For more general MSMs (e.g., logistic link, product terms of r_k and V)

Take observed data

id	R_0	R_1	R_2	R_3	R_4	R_5	age	Coll_grad	white	smoke	BMI	Y
1	1	1	0	0	0	1	30	1	1	0	25	1.02
2	0	1	0	0	1	1	35	0	1	0	27	-0.07
3	1	1	1	.	.	.	35	1	0	1	26	.
.	0	1	1	1	1	1	40	1	1	0	22	0.05
.	0	0	0	0	0	0	29	0	0	0	24	.
1584	1	1	0	0	1	1	39	0	1	0	21	-0.14

For more general MSMs (e.g., logistic link, product terms of r_k and V)

Only preserve values of baseline variables V

id	R_0	R_1	R_2	R_3	R_4	R_5	age	Coll_grad	white	smoke	BMI	Y
1							30	1	1	0	25	
2							35	0	1	0	27	
3							35	1	0	1	26	
.							40	1	1	0	22	
.							29	0	0	0	24	
1584							39	0	1	0	21	

For more general MSMs (e.g., logistic link, product terms of r_k and V)

Arm 1: set $\bar{r}_K$=$\bar{1}$ $(\text{i.e., set Rt to 1 for all time points})$

V

id	R_0	R_1	R_2	R_3	R_4	R_5	age	Coll_grad	white	smoke	BMI	Y
1	1	1	1	1	1	1	30	1	1	0	25	
2	1	1	1	1	1	1	35	0	1	0	27	
3	1	1	1	1	1	1	35	1	0	1	26	
.	1	1	1	1	1	1	40	1	1	0	22	
.	1	1	1	1	1	1	29	0	0	0	24	
1584	1	1	1	1	1	1	39	0	1	0	21	

For more general MSMs (e.g., logistic link, product terms of r_k and V)

Use MSM coefficients to predict $\mathrm{E}[Y^{\bar{r}_K=\bar{1},\bar{c}_{K+1}=0}|V_i]$

id	R_0	R_1	R_2	R_3	R_4	R_5	age	Coll_grad	white	smoke	BMI	Y
1	1	1	1	1	1	1	30	1	1	0	25	1.05
2	1	1	1	1	1	1	35	0	1	0	27	0.02
3	1	1	1	1	1	1	35	1	0	1	26	-0.05
.	1	1	1	1	1	1	40	1	1	0	22	0.07
.	1	1	1	1	1	1	29	0	0	0	24	1.21
1584	1	1	1	1	1	1	39	0	1	0	21	-0.20

For more general MSMs (e.g., logistic link, product terms of r_k and V)

Use MSM coefficients to predict $\mathrm{E}[Y^{\bar{r}_K=\bar{1},\bar{c}_{K+1}=0}|V_i]$

id	R_0	R_1	R_2	R_3	R_4	R_5	age	Coll_grad	white	smoke	BMI	Y
1	1	1	1	1	1	1	30	1	1	0	25	1.05
2	1	1	1	1	1	1	35	0	1	0	27	0.02
3	1	1	1	1	1	1	35	1	0	1	26	-0.05
.	1	1	1	1	1	1	40	1	1	0	22	0.07
.	1	1	1	1	1	1	29	0	0	0	24	1.21
1584	1	1	1	1	1	1	39	0	1	0	21	-0.20

Take average

For more general MSMs (e.g., logistic link, product terms of r_k and V)

Arm 2: set $\bar{r}_K = \bar{0}$ (i.e., set Rt to 0 for all time points)

Use MSM coefficients to predict $E[Y^{\bar{r}_K=\bar{0},\bar{c}_{K+1}=0}|V_i]$

id	R_0	R_1	R_2	R_3	R_4	R_5	age	Coll_grad	white	smoke	BMI	Y
1	0	0	0	0	0	0	30	1	1	0	25	1.11
2	0	0	0	0	0	0	35	0	1	0	27	0.20
3	0	0	0	0	0	0	35	1	0	1	26	-0.07
.	0	0	0	0	0	0	40	1	1	0	22	0.10
.	0	0	0	0	0	0	29	0	0	0	24	1.20
1584	0	0	0	0	0	0	39	0	1	0	21	-0.22

Take average