Causal inference based lifestyle coaching system for thyroid disease patients when lifestyle variables are continuous

[1]Yeongho Lee[†], Joonhyung Kim[2], Kyubo Shin[4], Youngin Kwon[1], Minhyun Kang[1], Sungil Kim[3], [1,3]Gi-Soo Kim[*],[5,6,7]Jae Hoon Moon[*],
[1]AIGS, UNIST,[2]LG Electronics ,
[3]Department of Industrial Engineering, UNIST,
[4]R&D division, THYROSCOPE INC.,[5]THYROSCOPE INC.,
[6]Department of Internal Medicine, SNU Bundang Hospital and SNU College of Medicine,
[7]Center for Artificial Intelligence in Healthcare, SNU Bundang Hospital

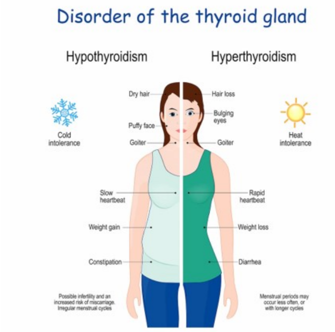
2023년 대한산업공학회 추계학술대회

Contents

Introduction

Research Background

Thyroid dysfunction is a common chronic disease that can be caused by either too much or too little secretion of thyroid hormones.



- As thyroid dysfunction is a chronic disease, it requires regular monitoring and treatment.
- Regular hospital visits for blood sampling and diagnosis increase the cost of medical care. This can be a burden to patients.

Introduction

Objective

- Daily habits may affect hormones and cause the symptoms.
- A lifestyle coaching system can help patients control their symptoms.
- Our approach: We provide coaching based on **causal inference**.

Example

Sleeping $\xrightarrow[\text{Causal Effect}]{\text{Found}}$ Nervousness

Daily Survey:
Sleeping 300
Training 40
Overwork 1
...

$\Longrightarrow$Recommend to increase sleep time

Method

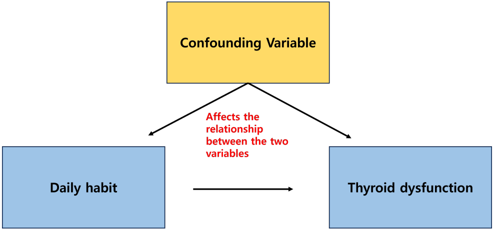
Causal Inference

We aim to estimate population average treatment effect(ATE),

$$\tau = \mathbb{E}[Y(t)] = \frac{1}{N}\sum_{i=1}^{N} Y_i(t).$$

$Y_i(t)$: Potential outcome of i-th patient when daily habit variable(treatment) $= t$.

Correlation $\neq$ Causation. Confounding variables should be adjusted for causal inference.



Assumptions

- Stable Unit Treatment Value Assumption(**SUTVA**)

  The outcome of one unit should be unaffected by the particular assignment of treatments to the other units.
- Strongly Ignorable(**Ignorability**)

  $$Y(t) \perp\!\!\!\perp T | X.$$

  T : realized treatment(daily habit value)
- Overlap

  $$0 < P(T = t | X = x) < 1.$$

Method

Generalized Propensity Score(GPS)

- (Binary treatment - Propensity Score) conditional probability of receiving a treatment given pre-treatment covarites X:

$$e(x) = \Pr(T = 1|X) = \mathbb{E}(T|X)$$

  where $X = (X_1, ..., X_p)$ is the collection of p covariates.
- (Continuous treatment - Generalized Propensity Score) the conditional density of the treatment given the covariates:

$$r(t, x) = f_{T|X}(t|x)$$

Method

Generalized Propensity Score

Property of Propensity Score

- Balancing property

$$X \perp\!\!\!\perp I(T=t)|r(t,x), \; X \perp\!\!\!\perp I(T=1)|r(t,x)$$

Method

Generalized Propensity Score

- By assumption **Ignorability**, $Y(t) \perp\!\!\!\perp T|X$.
- By property of GPS : $X \perp\!\!\!\perp I(T = t)|r(t, x)$,
- We can know that,

$$Y(t) \perp\!\!\!\perp I(T = t)|r(t, x)$$

- Estimation target

$$\mathbb{E}[Y(t)] = \frac{1}{N}\sum_{i=1}^{N} Y_i(t).$$

It can be calculated if we observe $Y_i(t)$ for all values t for all patients i. However for each patient i, we only observe $Y_i(T_i)$.

Method

Estimation of $E(Y(t))$ by Covariate Adjustment

$$
\begin{aligned}
\mathbb{E}[Y(t)] &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_i(t)|r(t,x_i)] \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_i(t)|T_i = t, r(t,x_i)](Y(t) \perp\!\!\!\perp I(T=t)|r(t,x)) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_i(T_i)|T_i = t, r(t,x_i)] \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_i|T_i = t, r(t,x_i)] \qquad (*)
\end{aligned}
$$

Generalized Propensity Score

Generalized Propensity Score

- Define Generalized Propensity Score,

$$f_\theta(T^*_{ij}|\mathbf{X^*_{ij}}) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left[-\frac{1}{2\hat{\sigma}^2}(T^*_{ij} - \mathbf{X^*_{ij}}^T\hat{\beta})^2\right]$$

(T^*_{ij} and $\mathbf{X^*_{ij}}$ are centered and normalized values of T_{ij} and $\mathbf{X}_{ij}$).

GPS Estimation

Covariate Balancing Generalized Propensity Score

- To estimate the parameters β and σ^2, we use the fact that when data are weighted by the inverse of GPS, the treatment and covariates should have covariance 0(Fong, Hazlett and Imai, 2018).

$$
\begin{aligned}
\mathbb{E}\left(\frac{f(T^*_{ij})}{f_\theta(T^*_{ij}|\mathbf{X^*_{ij}})}T^*_{ij}\mathbf{X^*_{ij}}\right) &= \int\left\{\int \frac{f(T^*_{ij})}{f_\theta(T^*_{ij}|\mathbf{X^*_{ij}})}T^*_{ij}\, dF(T^*_{ij}|\mathbf{X^*_{ij}})\right\}\mathbf{X^*_{ij}}dF(\mathbf{X^*_{ij}}) \\
&= \mathbb{E}(T^*_{ij})\mathbb{E}(\mathbf{X^*_{ij}}) = 0.
\end{aligned}
$$

- The covariate balancing generalized propensity score methodology increases the robustness to model misspecification by directly optimizing sample covariate balance between treatment and control groups.

GPS Estimation

Covariate Balancing Generalized Propensity Score

- By assumption,

$$f(T^*_{ij}) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{{T^*_{ij}}^2}{2}\right\}$$

- We can make weight function

$$\frac{f(T^*_{ij})}{f_\theta(T^*_{ij}|\mathbf{X^*_{ij}})} = \sigma exp\left[\frac{1}{2\sigma^2}(T^*_{ij} - {\mathbf{X^*_{ij}}}^T\beta)^2 - \frac{{T^*_{ij}}^2}{2}\right]$$

GPS Estimation

Proposed method : User Intercept

- GPS for a specific patient i,

$$f_\theta(T_{ij}|\mathbf{X}_{ij}, \text{patient id} = i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left[-\frac{1}{2\hat{\sigma}^2}(T_{ij} - \mathbf{X}_{ij}^T\hat{\beta} - \hat{\alpha}_i)^2\right]$$

- What is the probability of being a patient with a specific treatment and covariate?

$$\mathbb{P}(\text{patient id} = i|\mathbf{X}^*_{\mathbf{ij}}, T^*_{ij})$$

GPS Estimation

Proposed method : User Intercept

- We can find each user intercept using Multinomial Logistic Regression.

$$\mathbb{P}(\text{patient id} = i|\mathbf{X}^*_{\mathbf{ij}}, T^*_{ij}) = \frac{\mathbb{P}(T^*_{ij}|patient\ id = i, \mathbf{X}^*_{\mathbf{ij}})\mathbb{P}(patient\ id = i|\mathbf{X}^*_{\mathbf{ij}})}{\mathbb{P}(T^*_{ij}|\mathbf{X}^*_{\mathbf{ij}})}$$

$$\mathbb{P}(\text{patient id} = k|\mathbf{X}^*_{\mathbf{ij}}, T^*_{ij}) = \frac{\mathbb{P}(T^*_{ij}|patient\ id = k, \mathbf{X}^*_{\mathbf{ij}})\mathbb{P}(patient\ id = k|\mathbf{X}^*_{\mathbf{ij}})}{\mathbb{P}(T^*_{ij}|\mathbf{X}^*_{\mathbf{ij}})}$$

$$\frac{\mathbb{P}(\text{patient id} = i|\mathbf{X}^*_{\mathbf{ij}}, T^*_{ij})}{\mathbb{P}(\text{patient id} = k|\mathbf{X}^*_{\mathbf{ij}}, T^*_{ij})} = \frac{\mathbb{P}(T^*_{ij}|patient\ id = i, \mathbf{X}^*_{\mathbf{ij}})\mathbb{P}(patient\ id = i|\mathbf{X}^*_{\mathbf{ij}})}{\mathbb{P}(T^*_{ij}|patient\ id = k, \mathbf{X}^*_{\mathbf{ij}})\mathbb{P}(patient\ id = k|\mathbf{X}^*_{\mathbf{ij}})}$$

$$= \frac{\mathbb{P}(T^*_{ij}|patient\ id = i, \mathbf{X}^*_{\mathbf{ij}})}{\mathbb{P}(T^*_{ij}|patient\ id = k, \mathbf{X}^*_{\mathbf{ij}})} \times \frac{\mathbb{P}(patient\ id = i|\mathbf{X}^*_{\mathbf{ij}})}{\mathbb{P}(patient\ id = k|\mathbf{X}^*_{\mathbf{ij}})}$$

GPS Estimation

Proposed method : User Intercept

- Compare only treatment coefficients,

$$\frac{\mathbb{P}(\text{patient} = i|\mathbf{X}^*_{\mathbf{ij}}, T^*_{ij})}{\mathbb{P}(\text{patient} = k|\mathbf{X}^*_{\mathbf{ij}}, T^*_{ij})} = \begin{pmatrix} 1 & \mathbf{X}^*_{\mathbf{ij}} & T^*_{ij} \end{pmatrix}^T \vec{\gamma} = \gamma_{i0} + \mathbf{X}^*_{\mathbf{ij}}\gamma_{i1} + T^*_{ij}\gamma_{i2}$$

$$= \left[\frac{\frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left(-\frac{1}{2\hat{\sigma}^2} (T^*_{ij} - {\mathbf{X}^*_{\mathbf{ij}}}^T \hat{\beta} - \hat{\alpha}_i)^2 \right)}{\frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left(-\frac{1}{2\hat{\sigma}^2} (T^*_{ij} - {\mathbf{X}^*_{\mathbf{ij}}}^T \hat{\beta} - \hat{\alpha}_k)^2 \right)} \right]$$

$$= -\frac{1}{2\hat{\sigma}^2} \left\{ (2T^*_{ij} - 2{\mathbf{X}^*_{\mathbf{ij}}}^T \hat{\beta} - \alpha_i - \alpha_k)(-\alpha_i + \alpha_k) \right\}$$

$$T^*_{ij}\gamma_{i2} = -\frac{1}{2\hat{\sigma}^2}(2T^*_{ij})(-\alpha_i + \alpha_k)$$

$$\hat{\alpha}_i = \hat{\sigma}^2\gamma_{i2} + \alpha_k$$

GPS Estimation

Covariate Balancing Generalized Propensity Score

We find $\theta = (\beta, \sigma, \alpha_k)$ that makes the covariance between the treatment and covariate 0 when weighted by the inverse GPS.

$$\mathbf{F}_\theta = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\sigma exp\left[\frac{1}{2\sigma^2}(T^*_{ij} - \mathbf{X^*_{ij}}^T\beta - \hat{\alpha}_i)^2 - \frac{T^{*2}_{ij}}{2}\right] T^*_{ij}\mathbf{X^*_{ij}}\right)$$

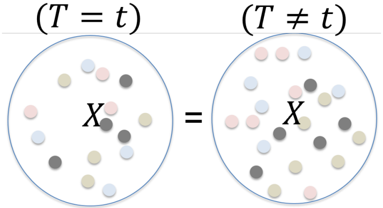
$$\hat{\theta} = \underset{\theta}{\text{argmin}}||\mathbf{F}_\theta||_2$$

This estimation procedure allows to find value of θ that satisfy the balancing property of GPS. When balancing property holds, the (*) equation is established, which allows causal inference.

Balance Test

Checking Covariate balance

- The property of GPS : $X \perp\!\!\!\perp I(T=t)|r(t,x)$



1. Divide treatment into Quantile bins.
2. Calculate the GPS according to the division, take the average, and divide it into deciles.
3. Divide the group by whether they belong to the Quantile.
4. T-test for Covariates.

Balance Test

Checking Covariate balance

- The Proportion of T-Tests with P-Values Greater Than or Equal to 0.1.
- If $p > 0.1$, GPS holds balancing property.

Balance Test(overall)		
Treatment	Traditional Method	Proposed Method
iodine	98.86%	98.86%
smoking	81.82%	89.77%
drinking	100%	98.48%
strength training	96.97%	96.97%
cardio training	96.59%	93.18%
sleep time	98.86%	96.59%
taken	100%	98.48%

Population Average Treatment Effect

Estimation of $E(Y(t))$ by Covariate Adjustment

- Model(*) is assumed to be a linear model with treatment and GPS values, interactions between the two values, and square terms as covariates.

$$\mathbb{E}[Y_{ij}|T_{ij}, r(T_{ij}, \mathbf{X}_{ij})]$$
$$= \begin{pmatrix}1 & r(T_{ij}, \mathbf{X}_{ij}) & T^*_{ij} & r(T_{ij}, \mathbf{X}_{ij})T^*_{ij} & r(T_{ij}, \mathbf{X}_{ij})^2 & T^{*2}_{ij}\end{pmatrix}^T \vec{\hat{\delta}}$$

- Using the estimated $\hat{\delta}$, estimate a specific symptom value

$$\mathbb{E}[\hat{Y(t)}] = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{E}[Y_{ij}|T_{ij} = t, r(t, \mathbf{X}_{ij})]$$
$$= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_i}\sum_{j=1}^{N_i}\begin{pmatrix}1 & r(T_{ij}, \mathbf{X}_{ij}) & T^*_{ij} & r(T_{ij}, \mathbf{X}_{ij})T^*_{ij} & r(T_{ij}, \mathbf{X}_{ij})^2 & T^{*2}_{ij}\end{pmatrix}$$

Experiment

Data Description

Data was gathered by Glandy™App of THYROSCOPE INC. from 150 hyperthyroidism patients for 15 months.

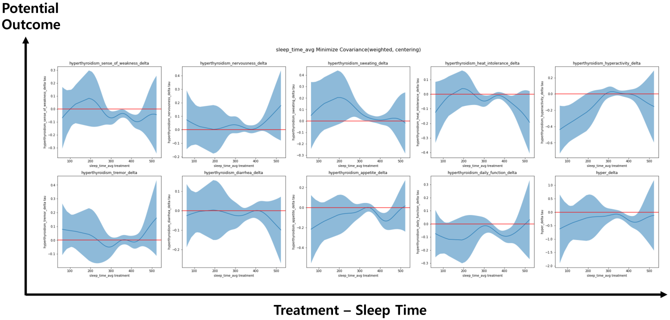
- X: covariate vector averaged over 28-day window
- Y: change in hyperthyroid symptom (average of 4th week−1st week)
- T: daily feature averaged over 28-day window

  (We moved the window with 7 days - stride.)

Type	Name of Feature	Description
Static Feature	birth	The year of the birth
	gender	Male or Female
	height	Height
	weight	Weight
Daily Feature	iodine_avg	The averaged intensity of intaking iodine
	drinking_avg	The averaged intensity of drinking alcohol
	smoking_avg	The averaged intensity of smoking cigarette
	st_training_avg	The averaged duration of strength exercise (min.)
	cd_training_avg	The averaged duration of aerobic exercise (min.)
	over_work_avg	The averaged status of doing overwork or not
	shift_work_avg	The averaged status of doing shift work or not
	sleep_time_avg	The averaged duration of sleep (min.)
	taken_avg	The averaged ratio of drug compliance
Hyperthyroidism Symptom	sense_of_weakness_delta	The changes for intensity of sense of weakness
	nervousness_delta	The changes for intensity of nervousness
	sweating_delta	The changes for intensity of sweating
	heat_intolerance_delta	The changes for intensity of sensitivity feeling heat
	hyperactivity_delta	The changes for intensity of hyperactivity
	tremor_delta	The changes for intensity of tremor
	diarrhea_delta	The changes for intensity of diarrhea
	appetite_delta	The changes for intensity of appetite
	daily_function_delta	The changes for intensity of function for daily life

Experiment

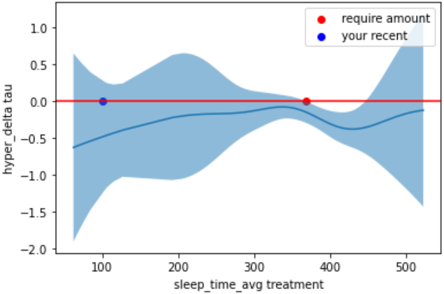
Result



- The confidence bands are obtained by bootstrapping.
- If $\hat{\mathbb{E}}[Y(t)]$ is negative on an interval of t, it means that symptoms are relieved when patients sleep for the amount of time in that interval.

Experiment

Recommendation Example



- We can provide a specific coaching to patients so as they reduce the thyroid dysfunction symptoms.
- If the patient is currently sleeping 100 minutes (blue dot), we can recommend him/her to sleep 260 minutes more since the confidence band starts to drop below the zero line at 360 minutes of sleep.

Acknowledgement

This work was supported by Institute of Information communications Technology Planning Evaluation(IITP) grant funded by the Korea government(MSIT) (No.2020-0-01336, Artificial Intelligence Graduate School Program(UNIST)) and the 2023 Research Fund (1.230019) of UNIST.