The Effect of Physical Activity on Sleep

Coco Kusiak, Hachem Saddiki, and Krishna Vemuri

12/12/2017

Contents

Step 0: Specify the Scientific Question

Background

50-70 million Americans have chronic sleep problems and 35.5% of adults report getting less than 7 hours of sleep a day (Disease Control and Prevention 2011). Everyone knows the implications of not getting enough sleep. It has been linked to many health problems such as obesity, mood disorders, and heart disease ("Consequences of Insufficient Sleep" 2017). Getting a good night's sleep gives the body the necessary time to recover and is shown to improve memory ("Sleep, Learning, and Memory" 2017). People across the country test herbal supplements, meditation and prescription drugs to fight insomnia. However, some studies have shown that exercise alone can help improve the quality of sleep (Foundation 2017). We set out to examine this relationship and see if exercise does indeed cause better sleep quality.

The Data

Our data come from the National Health and Nutrition Examination Survey conducted by the US National Center for Health Statistics and can found in the `NHANES` package in R. These data were collected between the years of 2009 and 2012 across the United States. It includes information on a variety of diseases, medical conditions, and health indicators. The original set has 76 variables and 10,000 observations ("About the National Health and Nutrition Examination Survey" 2012).

The `NHANES` includes duplicate rows in order to fix problems associated with oversampling. We decided to remove these duplicates. We also ran a series of stepwise regressions to identify key covariates. After these reductions, we have a final dataset with 10 variables and 4,654 observations. Unfortunately, although this dataset has information on multiple years, the participants all have unique identifiers and thus can not be analyzed in a longitudinal setting. `NA`s were transformed as factor levels labeled as "Undisclosed."

The Variables

Exposure

- `PhysActive`: is a binary variable ($Yes = 1$, $No = 0$) representing whether the participant does moderate or vigorous-intensity sports, fitness or recreational activities.

Outcome

- `SleepTrouble`: is a binary variable ($Yes = 1$, $No = 0$) representing whether the participant has reported having trouble sleeping to a healthcare professional.

Covariates

Demographic Covariates

1. `Gender`: is a binary variable ($Female = 0$, $Male = 1$) representing the gender of the participant.
2. `Age`: is a numeric variable representing the age in years at screening of study participant. Note that we do not consider participants younger than 20, and participants older than 80 were recorded as 80.
3. `Race1`: categorical variable representing the reported race of participant: `Mexican`, `Hispanic`, `White`, `Black`,`Other`.
4. `Education`: categorical variable representing the reported educational level of study participant: `8thGrade`, `9-11Grade`, `HighSchool`, `SomeCollege`, `CollegeGrad`.

5. `MaritalStatus`: categorical variable representing the reported marital status of study participant: `Married`, `Widowed`, `Divorced`, `Separated`, `NeverMarried`, `LivePartner`.
6. `HHIncomeMid`: numeric variable representing the median of reported income category of study participant.
7. `HomeOwn`: categorical variable representing whether the home of study participant is owned, rented or occupied by some other arrangement: `Home`, `Rent`, `Other`.

Health Covariates

8. `HealthGen`: categorical variable representing self-reported rating of participant's health in general: `Excellent`, `Vgood`, `Good`, `Fair`, `Poor`.
9. `DaysPhysHlthBad`: numeric variable representing the self-reported number of days participant's physical health was not good out of the past 30 days.
10. `DaysMentHlthBad`: numeric variable representing the self-reported number of days participant's mental health was not good out of the past 30 days.
11. `LittleInterest`: categorical variable representing the self-reported number of days participant had little interest in doing things: `None`, `Several`, `Majority`,`AlmostAll`.
12. `Depressed`: categorical variable representing the self-reported number of days participant felt down, depressed or hopeless: `None`, `Several`, `Majority`,`AlmostAll`.

Lifestyle Covariates

13. `TVHrsDay`: categorical variable representing the self-reported average number of hours per day participant watched TV over the past 30 days: `0_to_1hr`, `1_hr`, `2_hr`, `3_hr`, `4_hr`, `More_4_hr`.
14. `CompHrsDay`: categorical variable representing the self-reported average number of hours per day participant used a computer or gamind device over the past 30 days: `0_to_1hr`, `1_hr`, `2_hr`, `3_hr`, `4_hr`, `More_4_hr`.
15. `AlcoholDay`: numeric variable representing the reported average number of drinks consumed on days that participant drank alcohol beverages.
16. `SmokeNow`: binary variable ($Yes = 1$, $No = 0$) representing whether the participant currently smokes cigarettes regularly.
17. `RegularMarij`: binary variable ($Yes = 1$, $No = 0$) representing whether the participant has been or is a regular marijuana user (used at least once a month for a year).
18. `HardDrugs`: binary variable ($Yes = 1$, $No = 0$) representing whether the participant hs tried cocaine, crack cocaine, heroin or metamphetamine.

Scientific Question and Target Population

Among American adults aged 20 years or older, how does participating in moderate or vigorous-intensity sports, fitness or recreational activities affect sleep quality?

Step 1: Specify the Causal Model

Option 1

Our endogenous variables are X = {W, A, Y} which are defined as:

W= $\{W_d, W_h, W_l\}$
W_d = {Gender, Age, Race1, Education, MaritalStatus, HHIncomeMid, HomeOwn}
W_h = {HealthGen, DaysPhysHlthBad, DaysMentHlthBad, LittleInterest, Depressed}
W_l = {TVHrsDay, CompHrsDay, AlcoholDay, SmokeNow, RegularMarij, HardDrugs}

A = {PhysActive}
Y = {SleepTrouble}

Our exogenous variables are U = $U_{W_d}, U_{W_h}, U_{W_l}, U_A, U_Y$ P_U. These are the unmeasured factors which influence what values our endogenous variables, X can take.

Our structural causal model is defined as:

$W_d = f_{W_d}(U_{W_d})$
$W_h = f_{W_h}(W_d, U_{W_h})$
$W_l = f_{W_l}(W_d, W_h, U_{W_l})$
$A = f_A(W_d, W_h, W_l, U_A)$
$Y = f_Y(W_d, W_h, W_l, A, U_Y)$

A directed acyclic graph displays this SCM below.

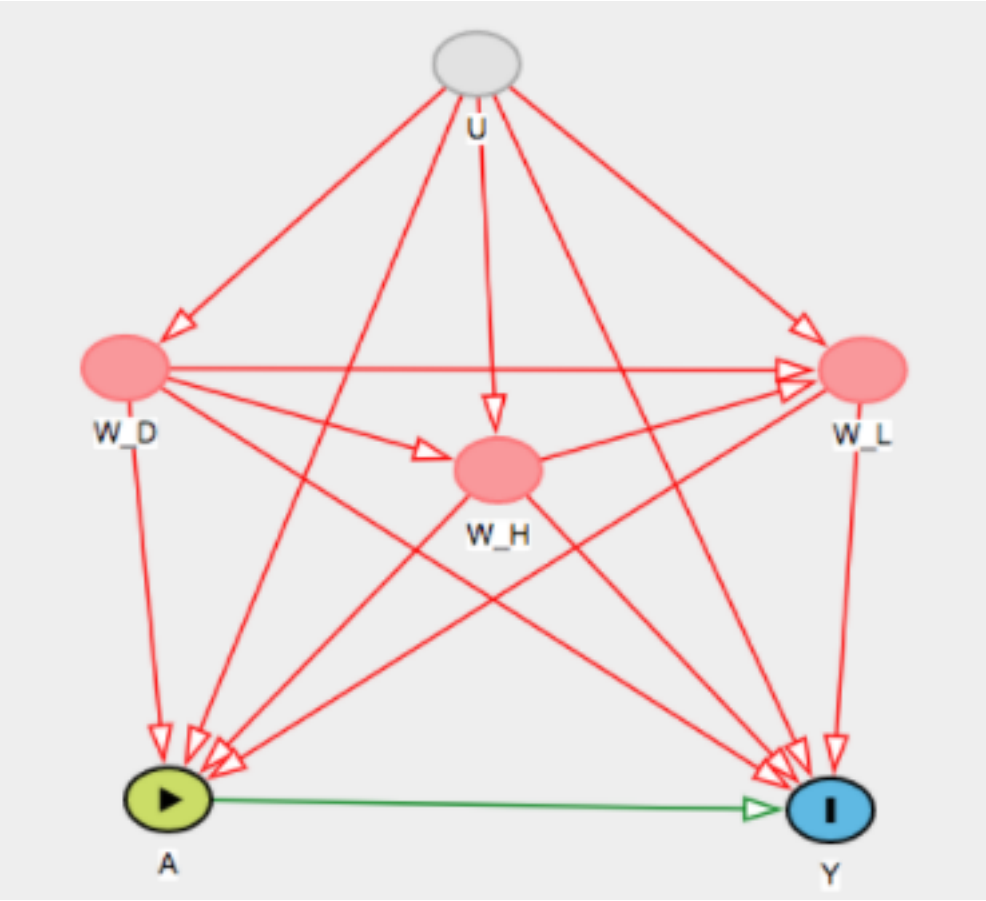

Figure 1: Structural Causal Model Option 1.

Option 2

Our endogenous variables X = {W, A, Z, Y} which are defined as:

W = $\{W_d, W_h, W_l\}$
W_d = {Gender, Age, Race1, Education, MaritalStatus, HHIncomeMid, HomeOwn}
W_h = {HealthGen, DaysPhysHlthBad, DaysMentHlthBad, LittleInterest}
W_l = {TVHrsDay, CompHrsDay, AlcoholDay, SmokeNow, RegularMarij, HardDrugs}

A = {PhysActive}
Z = {Depressed}
Y = {SleepTrouble}

Our exogenous variables are U = $\{U_{W_d}, U_{W_h}, U_{W_l}, U_A, U_Z U_Y\} \sim P_U$. These are the unmeasured factors which influence what values our endogenous variables, X can take.

Our structural causal model is defined as:

$W_d = f_{W_d}(U_{W_d})$
$W_h = f_{W_h}(W_d, U_{W_h})$
$W_l = f_{W_l}(W_d, W_h, U_{W_l})$
$A = f_A(W_d, W_h, W_l, U_A)$
$Z = f_Z(W_d, W_h, W_l, A, U_Z)$
$Y = f_Y(W_d, W_h, W_l, A, Z, U_Y)$

A directed acyclic graph displays this SCM below.

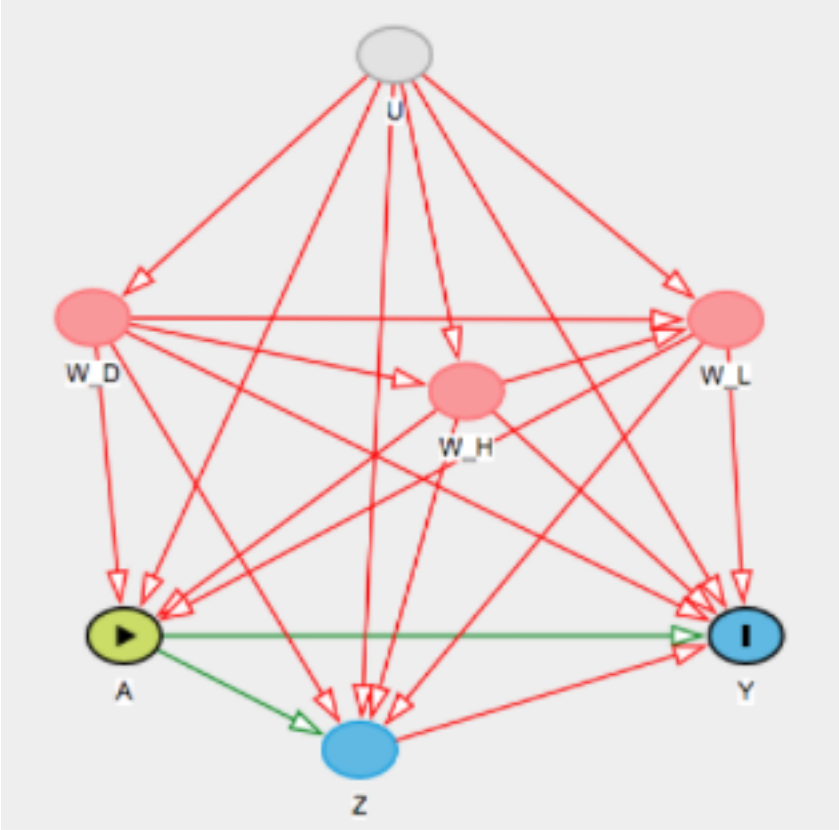

Figure 2: Structural Causal Model Option 2.

This alternative model defines Depressed as an intermediary between the exposure and outcome. Both models are valid, however we will proceed with the first model in the following analyses.

Step 2: Translation in the Target Causal Parameter Using Counterfactuals

Here, our intervention node A is physical activity. This is forcing each participant to either do moderate to vigorous physical activity or do no physical activity.

Our post-intervention SCM is then:

$W_d = f_{W_d}(U_{W_d})$
$W_h = f_{W_h}(W_d, U_{W_h})$
$W_l = f_{W_l}(W_d, W_h, U_{W_l})$
A = a
$Y_a = f_{W_l}(W_d, W_h, W_l, a, U_Y)$

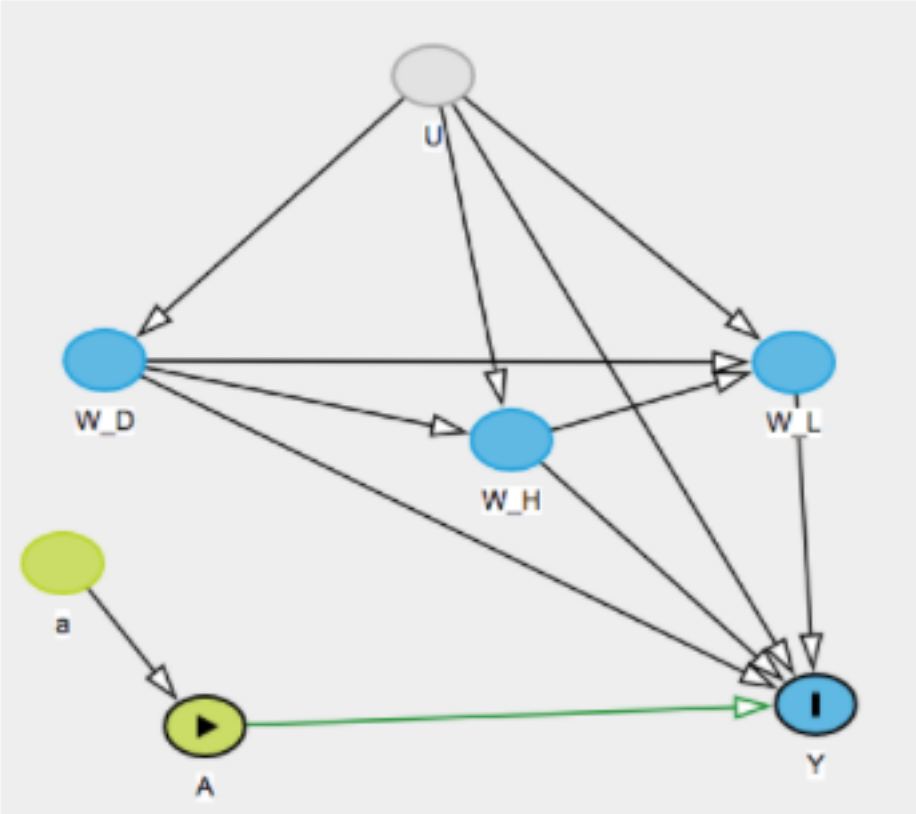

Figure 3: Post-Intervention Structural Causal Model.

The counterfactuals we are interested in are $Y_A : a \in A$ where A = {0,1}. Ya represents the individual's SleepTrouble if possibly contrary to fact, they received the intervention A = a.

Our causal parameter is the causal risk difference in SleepTrouble if every participant did moderate-vigorous physical activity vs. if no participant did physical activity. This can be written as:

$$\Psi^F(P_{U,X}) = P_{U,X}(Y_1 = 1) - P_{U,X}(Y_0 = 1) = E_{U,X}(Y_1) - E_{U,X}(Y_0)$$

Step 3: Specify the Observed Data and its Link

Our observed data consists of 4,654 independent and identically distributed copies of the random variable O which follows the distribution P_O. O $= \{W_l, W_d, W_h, A, Y\} \sim P_O$ where W_l consists of the lifestyle covariates, W_h consists of the health covariates, W_d consists of the demographic covariates, A is the intervention node of physical activity, and Y is the outcome variable representing sleep trouble. We can assume that our observed data comes from 4,654 i.i.d. samples from the data generating process described by M^F. This links our causal M^F with the statistical model M. This model M does not have any restrictions and is therefore non-parametric.

Step 4: Indentifiability

Step 5: The Statistical Model and Estimand

Step 6: Estimation

Step 7: Interpretation

References

"About the National Health and Nutrition Examination Survey." 2012. *National Center for Health Statistics*, september. doi:https://www.cdc.gov/nchs/nhanes/about_nhanes.htm.

"Consequences of Insufficient Sleep." 2017. *WGBH Educational FoundationDivision of Sleep Medicine at Harvard Medical School*. doi:http://healthysleep.med.harvard.edu/healthy/matters/benefits-of-sleep/sleep-and-disease-risk.

Disease Control, Centers for, and Prevention. 2011. "Unhealthy Sleep-Related Behaviors." *MMWR* 60: 233–34.

Foundation, National Sleep. 2017. "Study: Physical Activity Impacts Overall Quality of Sleep." doi:sleepfoundation.org/sleep-news/study-physical-activity-impacts-overall-quality-sleep.

"Sleep, Learning, and Memory." 2017. *WGBH Educational Foundation Division of Sleep Medicine at Harvard Medical School*. doi:http://healthysleep.med.harvard.edu/healthy/matters/benefits-of-sleep/learning-memory.