

# Causal Inference - Assignment Part1

Emilia Löscher

Kyuri Park

22 February, 2022

## 1. Draw a DAG representing a simple and fairly plausible causal system from your preferred topic of choice. Describe briefly the substantive motivation behind your DAG.

Many people with Obsessive-compulsive disorder (OCD) become depressed. According to Millet et al. (2004), the lifetime rates of major depression in OCD patients is about 81.2%. Several studies attempted to identify the causal mechanism in the comorbidity of OCD and depression (McNally et al. 2017; Zandberg et al. 2015) and speculated that OCD symptoms often precede and correspondingly activate the depression symptoms.

Given this background, here we proceed to look into the causal relationship between OCD and depression symptoms. We hypothesize that *distress associated with obsession* (OCD symptom) causes *feeling of guilt* (depression symptom) via several paths. Having been inspired by the DAG model from McNally et al. (2017), we construct our DAG with total 7 variables (nodes), which consist of OCD symptoms as well as depression symptoms. The specifics of the variables in our DAG is as follows:

< OCD symptoms >

- **ocdis**: distress caused by obsessions/compulsions
- **ocint**: interference due to obsessions/compulsions
- **occon**: difficulty controlling obsessions/compulsions

< Depression symptoms >

- **sad**: sadness
- **insom**: insomnia/sleeping problems
- **concen**: concentration/decision-making impairment
- **guilt**: guilt and self-blame

```
#varnames <- c("ocdis", "ocint", "occon", "sad", "insom", "concen", "guilt")
# Adj <- matrix(c(0,1,0,0,0,1,0,
#                0,0,1,1,0,0,0,
#                0,0,0,1,0,0,0,
#                0,0,0,0,1,0,1,
#                0,0,0,0,0,1,0,
#                0,0,0,0,0,0,1,
```

```

#           0,0,0,0,0,0,0), 7,7, byrow = TRUE, dimnames = list(varnames, varnames))
#
# qqgraph(Adj, labels = varnames)

ocddep <- dagify(
  ocdis ~ ocint,
  occon ~ ocdis,
  sad ~ ocdis + occon,
  insom ~ sad,
  concen ~ ocint + insom,
  guilt ~ concen + sad,
  exposure = "ocdis", # cause we are interested in
  outcome = "guilt", # effect variable we are interested in
  # set the coordites
  coords = list(x = c(ocdis = -1, ocint=-1, occon=-1, sad = 0, insom=0, concen=1, guilt = 1),
                y = c(ocdis = 0, ocint=1, occon=-1, sad = 0, insom=1, concen=1, guilt = 0))
)

ggdag_status(ocddep) + theme_dag()

```

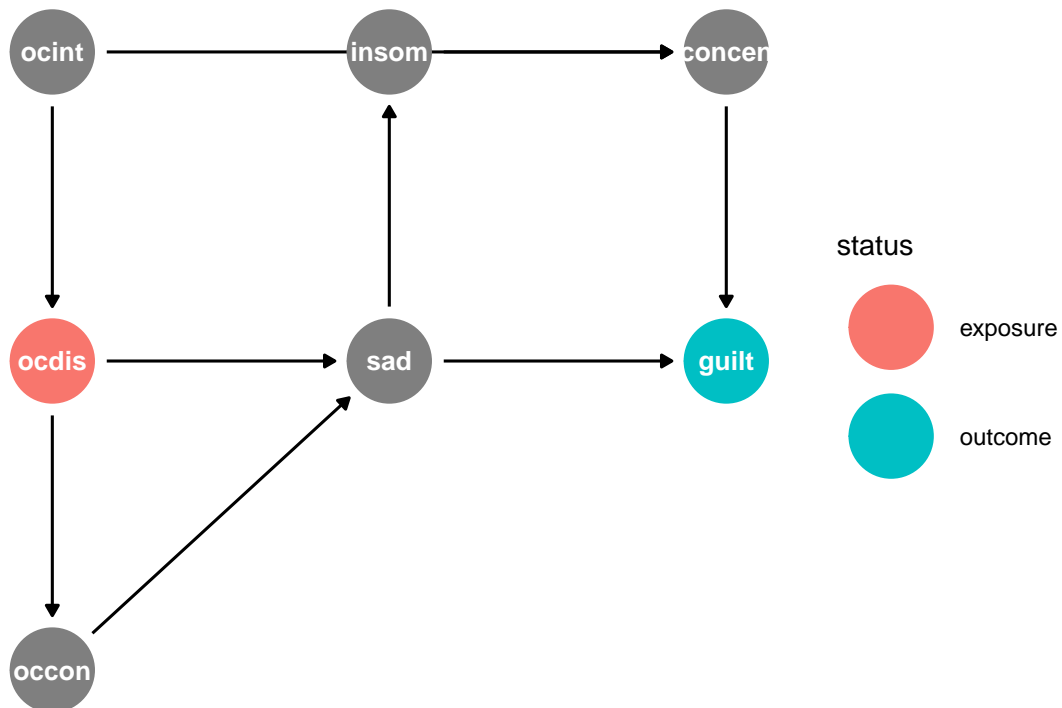


Figure 1: OCD-Depression DAG

## 2. Specify a structural causal model for your DAG.

In the following, the structural causal model for our DAG is specified.

$$\begin{aligned}ocint &:= \varepsilon_{ocint} \\ocdis &:= 3 \cdot ocint + \varepsilon_{ocdis} \\occon &:= 4 \cdot ocdis + \varepsilon_{occon} \\sad &:= 2 \cdot ocdis + 1 \cdot occon + \varepsilon_{sad} \\insom &:= 2 \cdot sad + \varepsilon_{insom} \\concen &:= 1 \cdot ocint + 4 \cdot insom + \varepsilon_{concen} \\guilt &:= 1 \cdot concen + 1 \cdot sad + \varepsilon_{guilt}\end{aligned}$$

where  $\varepsilon_{ocint}, \dots, \varepsilon_{guilt} \sim N(0, SD_i)$  with  $i \in ocint, \dots, guilt$ .

The standard deviations ( $SD_i$  for  $i \in ocint, \dots, guilt$ ) are taken from Table 1 on page 1206 in the paper. The same holds for the mean in the errors for the variable *ocint*.

We came up with the intercept and regression coefficients ourselves. To determine them, we took the DAG and the thickness of the arrows on page 1211 into account and adjusted them to the DAG we chose.

## 3. Set a seed in R. Generate data from your SCM with a sample size of 500 units

```
## With mean based on the data
# Setting a seed
set.seed(1705)

# Defining the sample size
n <- 1500

#Generating Data
#with means
ocint <- rnorm(n, 2.69, 0.82)
ocdis <- 3.5 * ocint + rnorm(n, 2.81, 0.76)
occon <- 3 * ocdis + rnorm(n, 2.67, 0.76)
sad <- 2.3 * ocdis + 1 * occon + rnorm(n, 1.55, 0.94)
insom <- 2 * sad + rnorm(n, 0.81, 1.07)
concen <- 1 * ocint + 4 * insom + rnorm(n, 1.48, 0.87)
guilt <- 1.5 * concen + 3.3 * sad + rnorm(n, 1.56, 1.17)
```

```

#Creating the data set
OCDDEP<- data.frame(ocint, ocdis, occon, sad, insom, concen, guilt)

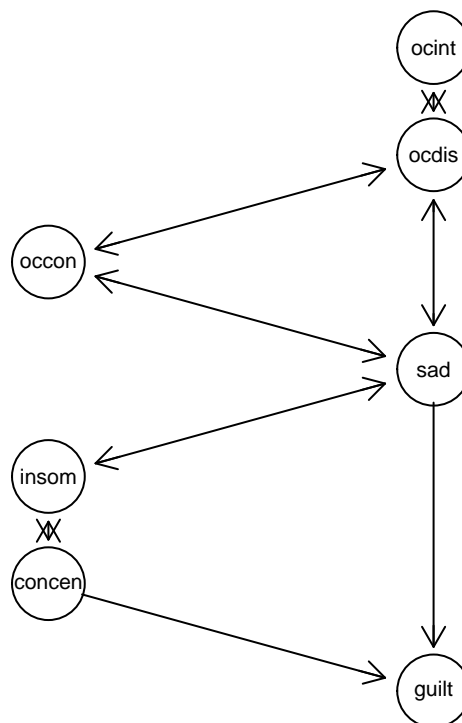
suffStat <- list(C= cor(OCDDEP), n = nrow(OCDDEP))

varnames <- colnames(OCDDEP)
pc_fit <- pc(suffStat = suffStat, indepTest = gaussCIttest, alpha = 0.01,
            labels = varnames)

# plot the estimated CPDAG
if (require(Rgraphviz)) {
  plot(pc_fit, main = "Inferred CPDAG using PC algorithm")
}

```

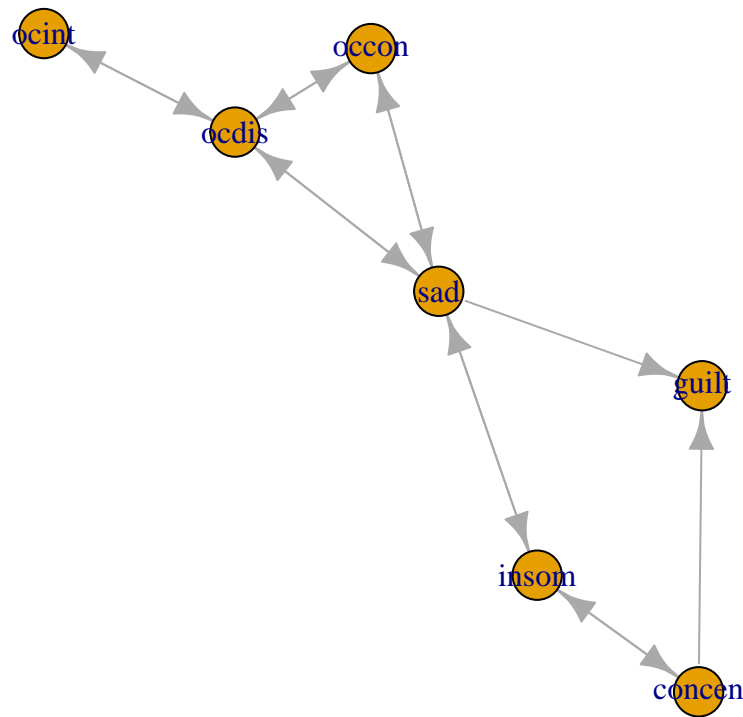
### Inferred CPDAG using PC algorithm



```

# better with igraph?
library(igraph)
iplotPC(pc_fit)

```



```

## With mean = 0

# Setting a seed
set.seed(1705)

# Defining the sample size
n <- 1500

#Generating the data
#with means = 0
ocint <- rnorm(n, 0, 0.82)
ocdis <- 3 * ocint + rnorm(n, 0, 0.76)
occon <- 4 * ocdis + rnorm(n, 0, 0.76)
sad <- 2 * ocdis + 1 * occon + rnorm(n, 0, 0.94)
insom <- 2 * sad + rnorm(n, 0, 1.07)
concen <- 1 * ocint + 4 * insom + rnorm(n, 0, 0.87)
guilt <- 1 * concen + 1 * sad + rnorm(n, 0, 1.17)

#Creating the data set
OCDDEP<- data.frame(ocint, ocdis, occon, sad, insom, concen, guilt)

```

```

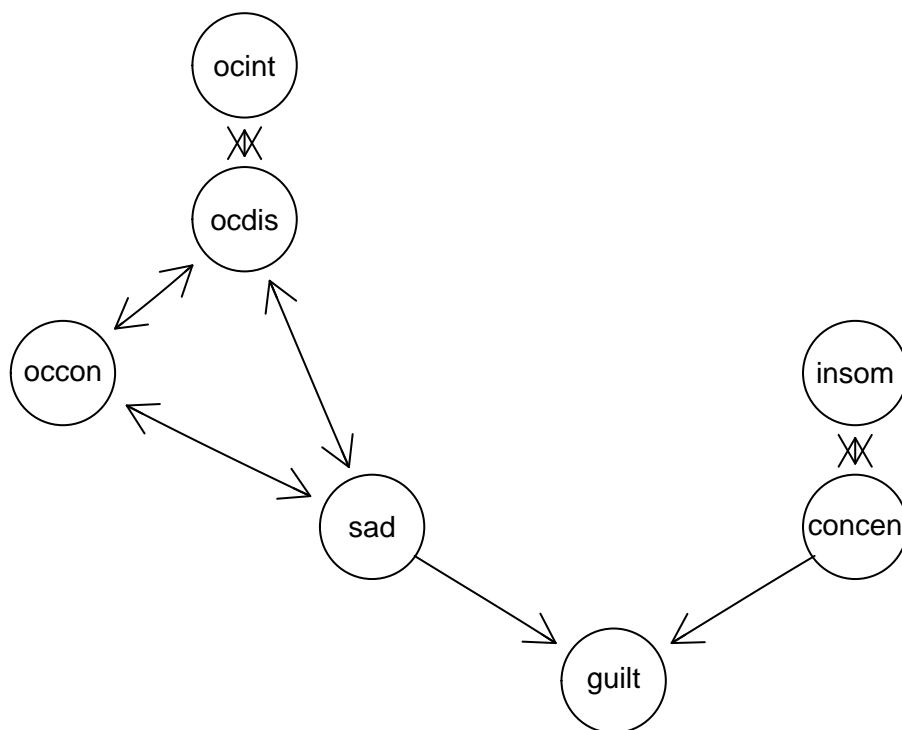
suffStat <- list(C= cor(OCDDEP), n = nrow(OCDDEP))

varnames <- colnames(OCDDEP)
pc_fit <- pc(suffStat = suffStat, indepTest = gaussCIttest, alpha = 0.01,
            labels = varnames)

# plot the estimated CPDAG
if (require(Rgraphviz)) {
  plot(pc_fit, main = "Inferred CPDAG using PC algorithm")
}

```

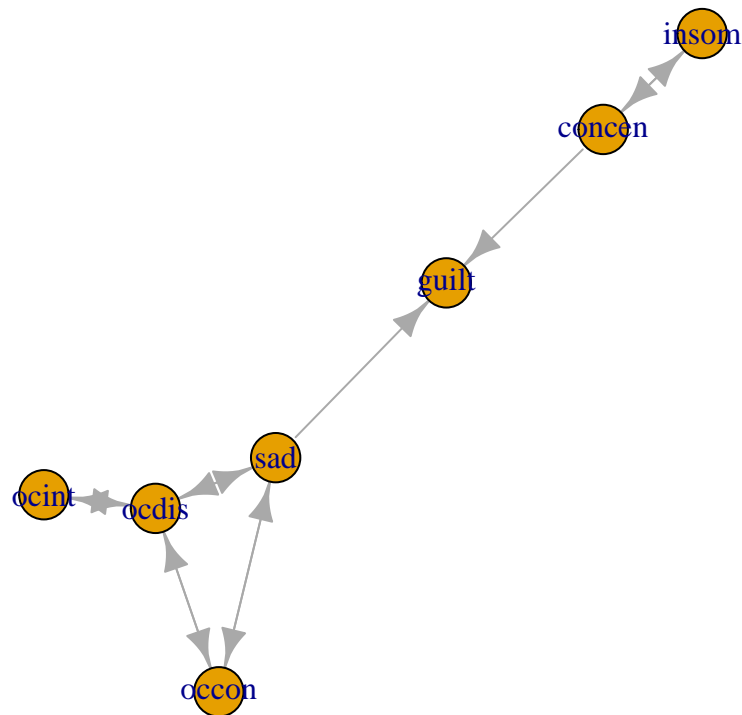
### Inferred CPDAG using PC algorithm



```

# better with igraph?
library(igraph)
iplotPC(pc_fit)

```



4. Use the PC-algorithm on the generated data to ‘discover’ the structure of your causal system

- a. Is your true DAG covered in the Markov equivalence class provided by the algorithm? Estimate the CPDAG

```

# suffStat_OCD <- list(C = cor(OCDdata), n = nrow(OCDdata))
#
# pc_OCD <- pc(suffStat = suffStat_OCD, indepTest = gaussCItest, p = ncol(OCDdata), alpha = 0.
#
# plot(pc_OCD, main = " Inferred CPDAG using pcalg")

```

- b. Provide the CP-DAG. To what extent did the procedure correctly recover which relationship were absent/present/directed?

5. data. Choose two variables for which they should estimate the causal relationship

- a. Cause variable: *ocdis*
- b. Outcome variable: *guilt*

## 6. For the causal effect specified in 5

- a. What is the true causal effect of the cause on the outcome variable based on your SCM?

```
1 * 2 + 1 * 4 * 2 * 2 + 1 * 1 * 4
```

```
## [1] 22
```

22?

- b. Based on the true DAG, what linear regression model should be used to estimate the causal effect correctly?

```
# mod_OCD <- lm(guilt ~ OCDdist + OCDint, data = OCDdata)
# summary(mod_OCD)
```

?

- c. Estimate the causal effect with this regression model based on your generated data. To what extent is the true effect recovered?

```
# cpdag_mat_OCD <- as(pc_OCD,"matrix")
#
# # Each row is a DAG adjacency matrix in vector form (by rows)
# res_OCD <- pdag2allDags(cpdag_mat_OCD)
#
# # We can get the adjacency matrix of an individual DAG using
# resOCD_dags <- list()
# for(i in 1:nrow(res_OCD$dags)){
#   resOCD_dags[[i]] <- t(matrix(res_OCD$dags[i,],7,7,byrow = TRUE))
# }
#
#
# ida(2,7, cov(OCDdata), pc_OCD@graph, verbose = TRUE)
```

The third CP-DAG does a good job?!

```
# res_Adj <- resOCD_dags[[3]]
#
# qgraph(res_Adj)
```



## 7. We will need a dichotomous cause variable

- a. Make a dichotomized version of the cause variable from question 5&6 in your dataset, for example by assigning scores lower than the mean a 0 and the rest a 1.

```
#Dichtomizing the effect variable OCDdist
# mean_OCDdist <- mean(OCDdistress)
# OCDdist_d <- rep(1, length(OCDdistress))
# OCDdist_d[which(OCDdistress < mean_OCDdist)] <- 0
#
# OCDdata <- cbind(OCDdata, OCDdist_d)
```

- b. Use the correct model for the causal effect (the one from 6b) to estimate the causal effect again, but now with your dichotomized cause variable. Discuss the results.

```
# mod_OCD_d <- lm(guilt ~ OCDdist_d + OCDint, data = OCDdata)
# summary(mod_OCD_d)
```

?

## 8. Prepare the data for the second causal inference assignment

- a. Create a data frame with all your variables

```
# save(OCDdata, file = "LoscherPark_Assign1_Data.RData")
```

- b. Make a brief .txt file See LoscherPark\_Assign1.txt.

## Reference

McNally, R.J., Mair, P., Mugno, B.L., & Riemann, B.C. (2017). Co-Morbid Obsessive–Compulsive Disorder and Depression: A Bayesian Network Approach. *Psychological Medicine*. 47(7): 1204–14. <https://doi.org/10.1017/S0033291716003287>.

Millet, B., Kochman, F., Gallarda, T., Krebs, M.O., Demonfaucon, F., Barrot, I., Bourdel, M.C., Olie, J.P., Loo, H., & Hantouche, E.G. (2004). Phenomenological and Comorbid Features Associated in Obsessive–Compulsive Disorder: Influence of Age of Onset. *Journal of Affective Disorders*. 79(1-3): 241–46.

Zandberg, L.J., Zang, Y., McLean, C.P., Yeh, R., Simpson, H.B., & Foa, E.B. (2015). Change in Obsessive-Compulsive Symptoms Mediates Subsequent Change in Depressive Symptoms During Exposure and Response Prevention. *Behaviour Research and Therapy*. 68:76–81.