

CA1 van Opstal

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(foreign)
library(tidyverse)
```

```
## -- Attaching packages -----

## v ggplot2 3.2.1    v purrr  0.3.2
## v tibble  2.1.3    v dplyr  0.8.3
## v tidyr   0.8.3    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0

## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dagitty)
library(ggdag)
```

```
##
## Attaching package: 'ggdag'

## The following object is masked from 'package:ggplot2':
##
##   expand_scale

## The following object is masked from 'package:stats':
##
##   filter
```

```
library(dplyr)
library(tinytex)
```

II

```
library(haven)
nhis <- read_dta("nhis2009_subsample_v1112.dta")
```

II

- a) Two potential outcomes:
- b) It could be that the insurance allows the person to get drugs and treatment and thus the insurance could increase health. It is also possible that the person will be less careful with their health because they are insured. Therefore an insurance could decrease the health

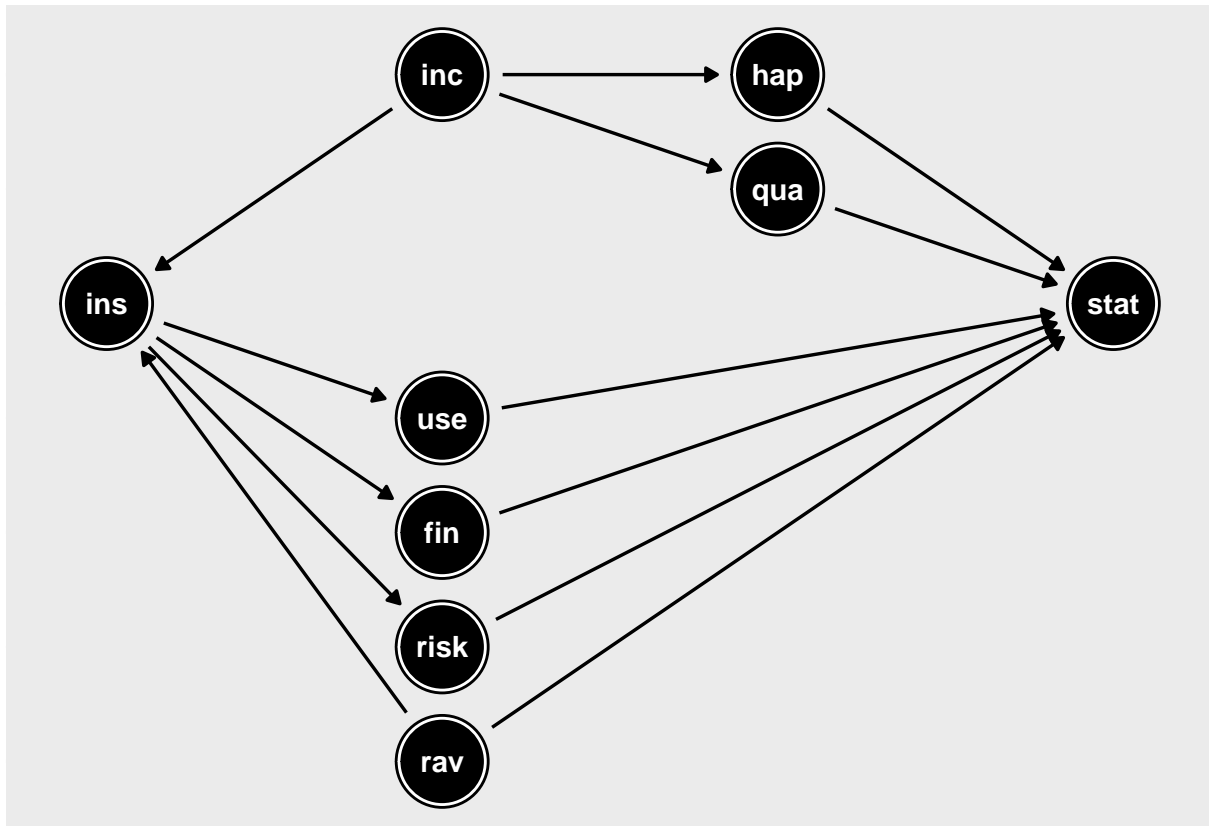
IV

```
ca1a <- dagitty('dag{
  ins [pos="0,2"]
  inc [pos="1,4"]
  qua [pos="2,3"]
  hap [pos="2,4"]
  stat [pos="3,2"]
  use [pos="1,1"]
  fin [pos="1,0"]
  risk [pos="1,-1"]
  rav [pos="1,-2"]

  ins <- inc -> hap -> stat
  inst -> stat
  ins -> use -> stat
  ins -> fin -> stat
  ins <- inc -> qua -> stat
  ins -> risk -> stat
  ins <-rav -> stat
}')
ggdag(ca1a)
```

```
## Warning: Removed 1 rows containing missing values (geom_dag_node).
```

```
## Warning: Removed 1 rows containing missing values (geom_dag_text).
```



- a) People with a high income are more likely to have a health-insurance aswell as a better health status through the better quality of food they eat. Therefore, not including income will overestimate the effect of health insurance since it doesn't take into account the better health status through income.
- b) The same goes for the path health insurance <- income -> happiness -> health status as in question a. People who have higher income are still more likely to have an health-insurance and they are more likely to be happy and thus more likely to have a better health-status. Therefore, not including the happiness path would overestimate the effect of health-insurance on health.
- c) It is not a case of selection bias since the risky behavior, in this question, does not affect whether or not you have a health insurance. It only affects health status and is affected by health insurance, therefore, it is a mediator
- d) Risk aversion would cause a selection bias since it would affect whether or not someone buys a health-insurance and it affects the health status of someone. (Someone who is more risk averse will buy an insurance and behave more safely).

V

```

sumhlth <- nhis %>%
  filter(fml==0) %>%
  group_by(hi) %>%
  summarise(mean_hlth = weighted.mean(hlth,perweight))
sumhlth

```

```
## # A tibble: 2 x 2
##       hi mean_hlth
##   <dbl>   <dbl>
## 1     0     3.70
## 2     1     4.01
```

```
suminc <- nhis %>%
  filter(fml==0) %>%
  group_by(hi) %>%
  summarise(mean_inc = weighted.mean(incmp,perweight))
suminc
```

```
## # A tibble: 2 x 2
##       hi mean_inc
##   <dbl>   <dbl>
## 1     0 44234.
## 2     1 98921.
```

VI

```
summary(lm(hlth ~ hi, data=nhis, subset= fml==0, weights = perweight))
```

```
##
## Call:
## lm(formula = hlth ~ hi, data = nhis, subset = fml == 0, weights = perweight)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -347.42  -41.47   -0.47   49.69  175.89
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.69565    0.02628  140.65  <2e-16 ***
## hi           0.31325    0.02827   11.08  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.68 on 9393 degrees of freedom
## Multiple R-squared:  0.0129, Adjusted R-squared:  0.01279
## F-statistic: 122.7 on 1 and 9393 DF,  p-value: < 2.2e-16
```

These results suggest an 0.3 increase on health status (on a scale of 1 to 5) for husbands who do own an health insurance in comparison to those who do not own an health-insurance.

```
summary(lm(hlth ~ hi + incmp, data=nhis, subset= fml==0, weights = perweight))
```

```
##
## Call:
## lm(formula = hlth ~ hi + incmp, data = nhis, subset = fml ==
##      0, weights = perweight)
```

```
##
## Weighted Residuals:
##      Min      1Q   Median      3Q      Max
## -304.749  -36.118    2.906   44.129  194.547
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.495e+00  2.734e-02 127.831  <2e-16 ***
## hi          6.502e-02  2.997e-02   2.169   0.0301 *
## incmp       4.539e-06  2.131e-07  21.303  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.36 on 9392 degrees of freedom
## Multiple R-squared:  0.0584, Adjusted R-squared:  0.0582
## F-statistic: 291.2 on 2 and 9392 DF,  p-value: < 2.2e-16
```

These results suggest that having a health-insurance increase the health status of an husband by 0.06 on a scale of 1 to 5 compared to an husband without health-insurance. Income has diminished a lot of the effect of health insurance compared to the regression based solely on whether someone does or does not have an health-insurance.