

Directed Acyclic Graphs

Karen Leffondré
University of Bordeaux
Inserm U1219 Bordeaux Population Health

Atelier Inserm
Best practices and recent advances in causal analyses
Practical phase

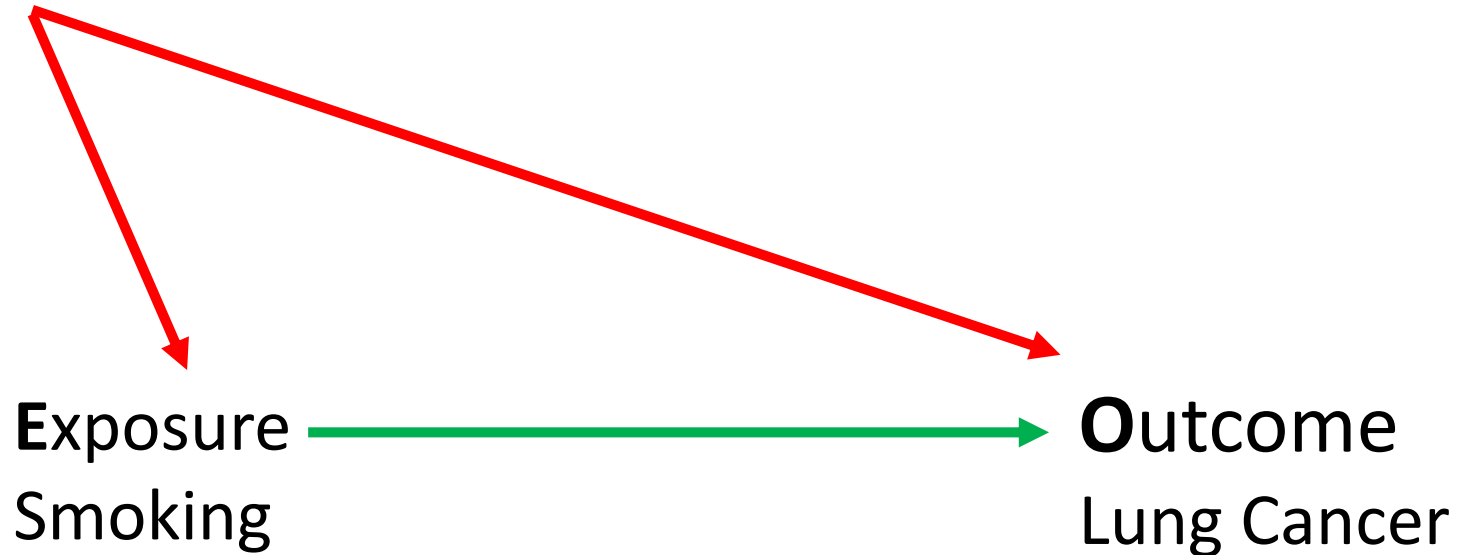
October 13-16, 2025 - Poitiers

DAGs in epidemiology

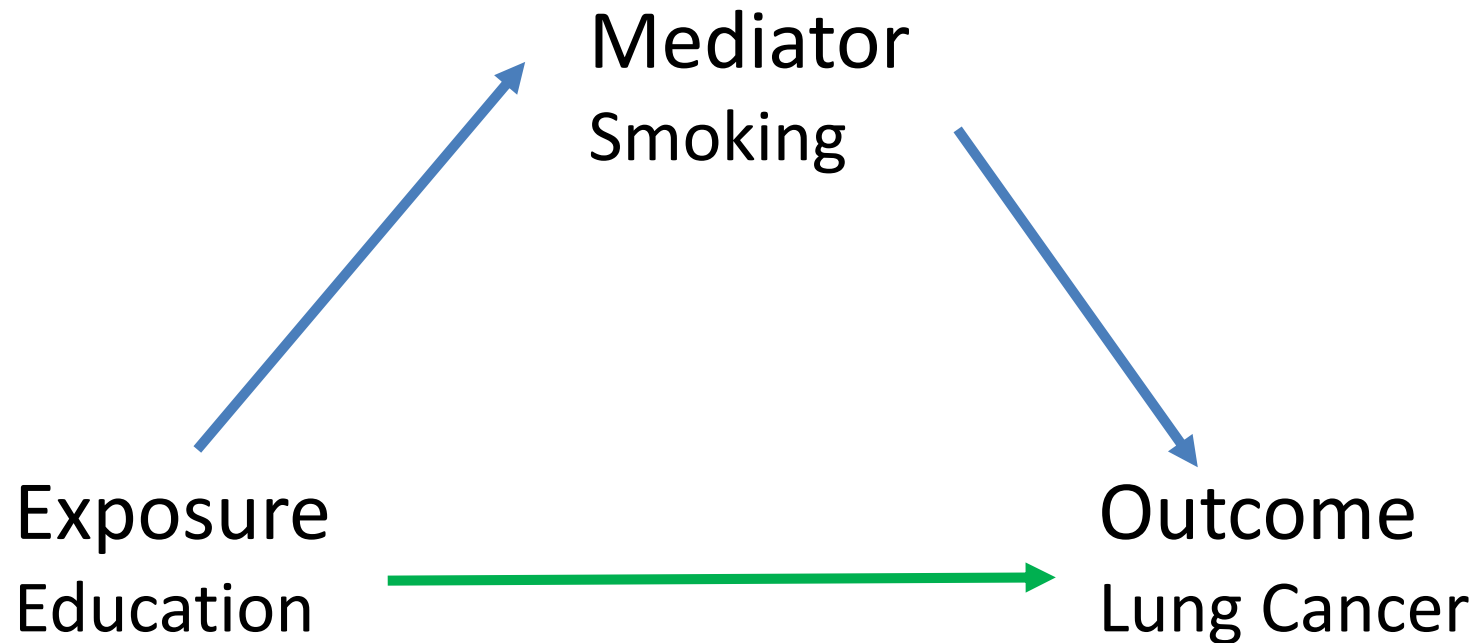
- DAGs are used in epidemiology to represent **hypothetical** causal relationships among variables of interest
 - At the time of developing the study design
 - To identify potential selection and confounding bias
 - Before etiological analysis
 - To identify the set of covariates for adjustment
- Pearl. [Causal diagrams for empirical research](#). *Biometrika* 1995
- Greenland, Pearl, Robins. [Causal diagrams for epidemiological research](#). *Epidemiol* 1999
- Glymour, Greenland. [Causal diagrams](#). In: Rothman, Greenland, Lash. *Modern Epidemiology* 2008
- Digitale, Martin, Glymour. [Tutorial on DAG](#). *J Clin Epidemiol* 2022

Confounder

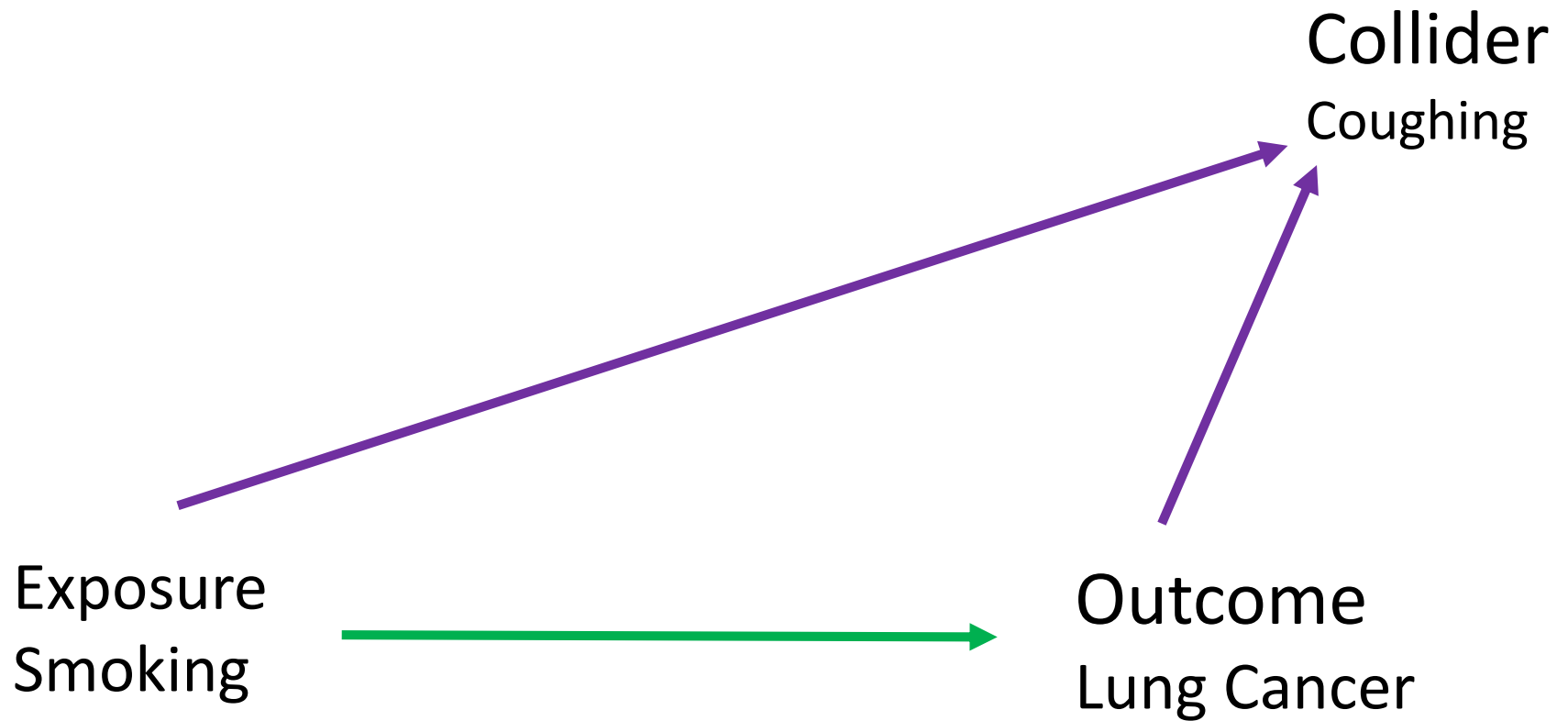
Confounder
Education



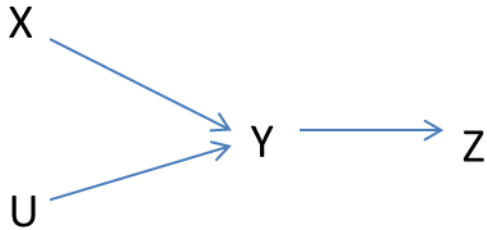
Mediator



Collider

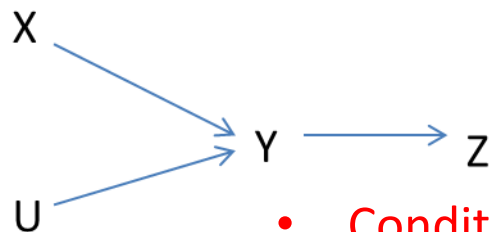


Marginal associations in a DAG



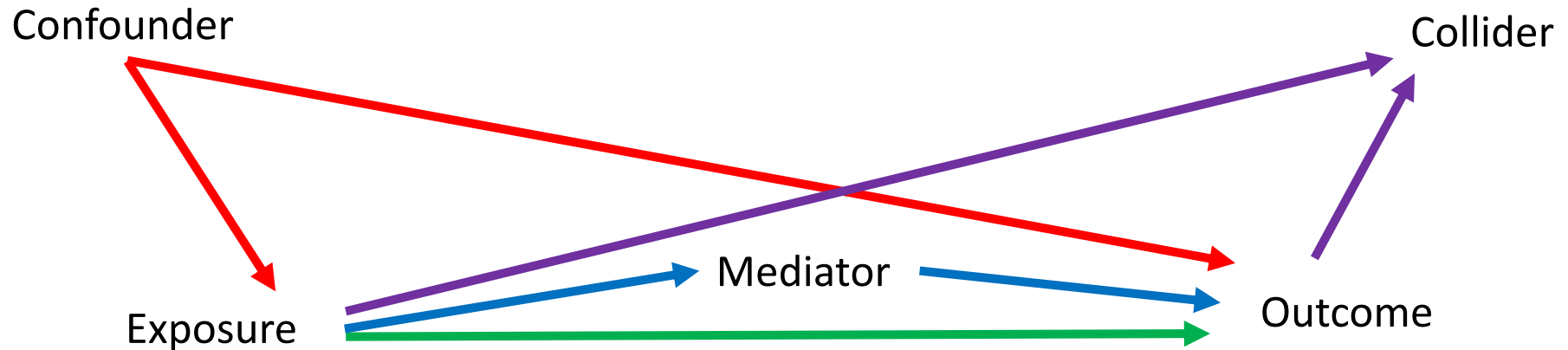
- A path is said to be **open** (unblocked, active) if there is no collider on the path
 - Path $X \rightarrow Y \rightarrow Z$
 - Path $U \rightarrow Y \rightarrow Z$
- A path is said to be **closed** (blocked, inactive) if there is a collider on the path
 - Path $U \rightarrow Y \leftarrow X$
 - Path $X \rightarrow Y \leftarrow U$
- If there is no open path between U and X, they are **marginally independent**
 - $\Pr(X = x) = \Pr(X = x \mid U = u)$

Conditional associations in a DAG



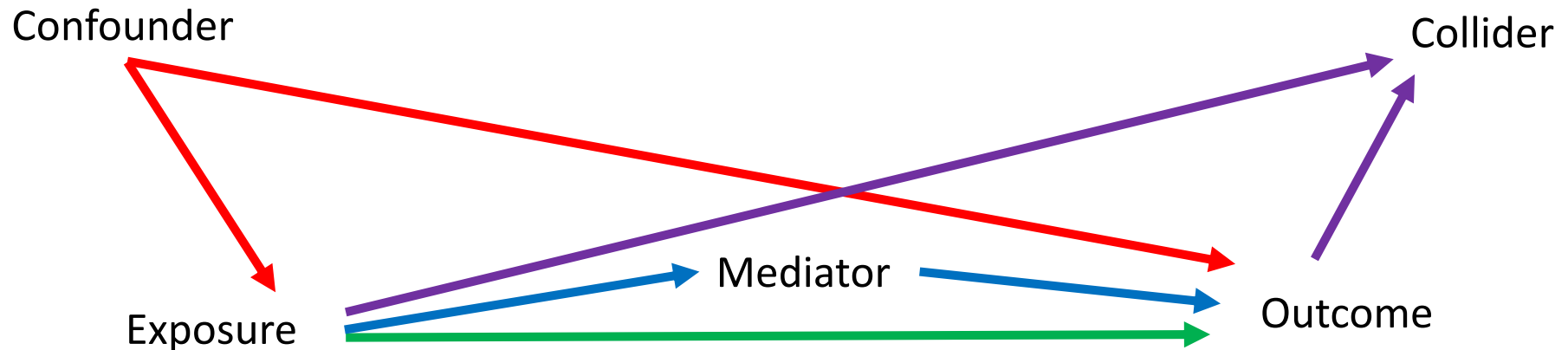
- **Conditioning on a noncollider closes the path**
 - Conditioning on Y close the path $X \rightarrow Y \rightarrow Z$
 - X and Z are marginally associated
but after conditioning on Y, they become independent
 $\Pr(Z = z \mid Y = y) = \Pr(Z = z \mid Y = y, X = x)$
➤ X and Z are **independent conditionally** on Y (Y separates Z from X)
- **Conditioning on a collider opens the path**
 - Conditioning on Y open the path $X \rightarrow Y \leftarrow U$
 - X and U are marginally independent
 $\Pr(X = x) = \Pr(X = x \mid U = u)$
but after conditioning on Y, they become associated
➤ X and U are **associated conditionally** on Y

Identification of the set of covariates for adjustment



- Identify all paths from the exposure to the outcome
 - Causal paths (all paths where all arrows go from the exposure to the outcome)
 - Not causal paths (all other paths: confounding paths, paths with a collider)
- Causal paths
 - Naturally open, we don't want to close them
- Confounding paths (backdoor paths)
 - Naturally open, we want to close them
(because association flows along all open paths)

Identification of the set of covariates for adjustment

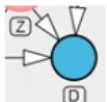


- **Paths with a collider**
 - Naturally closed, we don't want to open them (because association cannot flow along closed path)
- **Total effect**
 - Effect of the exposure on the outcome through all causal paths
- **Sufficient set of covariates for adjustment**
 - After adjustment, all causal paths are open and all non-causal paths are closed


Softwares

- The daggle app
 - Hanly et al. [The daggle app - a tool to support learning and teaching the graphical rules of selecting adjustment variables using DAGs](#). *Int J Epidemiol* 2023
 - DAGitty
 - Textor, Hardt, Knüppel. DAGitty: [A graphical tool for analyzing causal diagrams](#). *Epidemiol* 2011.
 - Textor et al. [Robust causal inference using DAGs: the R package 'dagitty'](#). *Int J Epidemiol* 2016
- « **Just** » have to put your DAG, and it gives you the set of covariates for adjustment


DAGitty website




Launch DAGitty online in your browser.



Download DAGitty's source for offline use.



Learn more about DAGs and DAGitty.



The R package "dagitty" is available on [CRAN](#) or [github](#).

DAGitty is developed and maintained by [Johannes Textor](#) ([Institute for Computing and Information Sciences](#), [Radboud University](#), and Medical BioSciences department, [Radboudumc](#), Nijmegen, The Netherlands).

Many algorithms implemented in DAGitty were developed in close collaboration with [Maciej Liśkiewicz](#) and [Benito van der Zander](#), University of Lübeck, Germany (see literature references below).

DAGitty development happens on [github](#). You can download all source code from there and also get involved.

How can I get help?



Johannes Textor
[@johannes_textor@mastodo...](#)

After 12 years, I am moving the [#dagitty](#) website to a new server. There will be interruptions and issues in the coming days. I hope everything will be up again smoothly soon. Thanks for your patience!

08 oct. 2023, 11:20 ·  · Web

0 boost · 5 favoris

Changelog

2023-10-07

Moved to a new webserver after 12 years.

2023-07-11

Version 3.1 is out, featuring selection variables.

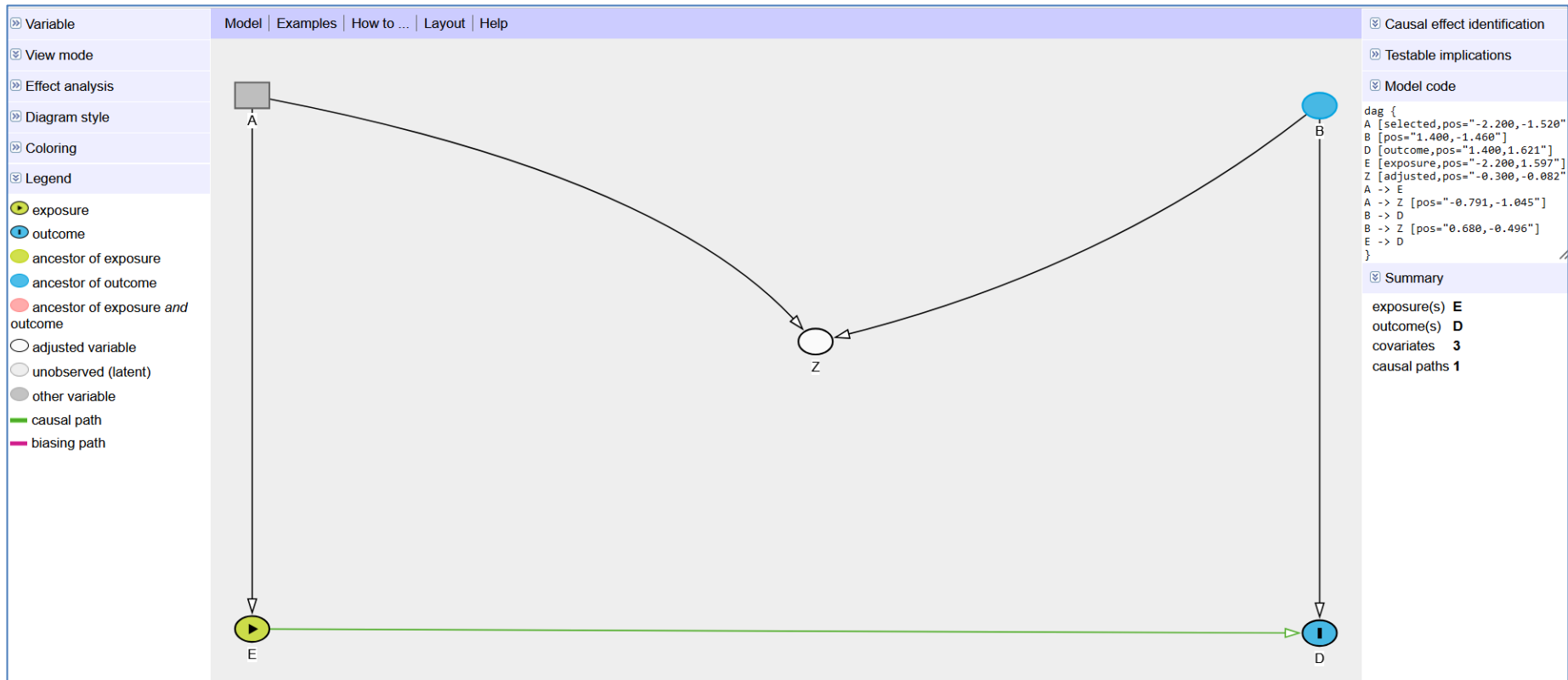
2020-01-09

Version 3.0 has been released! Complete reimplementaion of the interface, should work with mobile/touch now.

2018-04-04

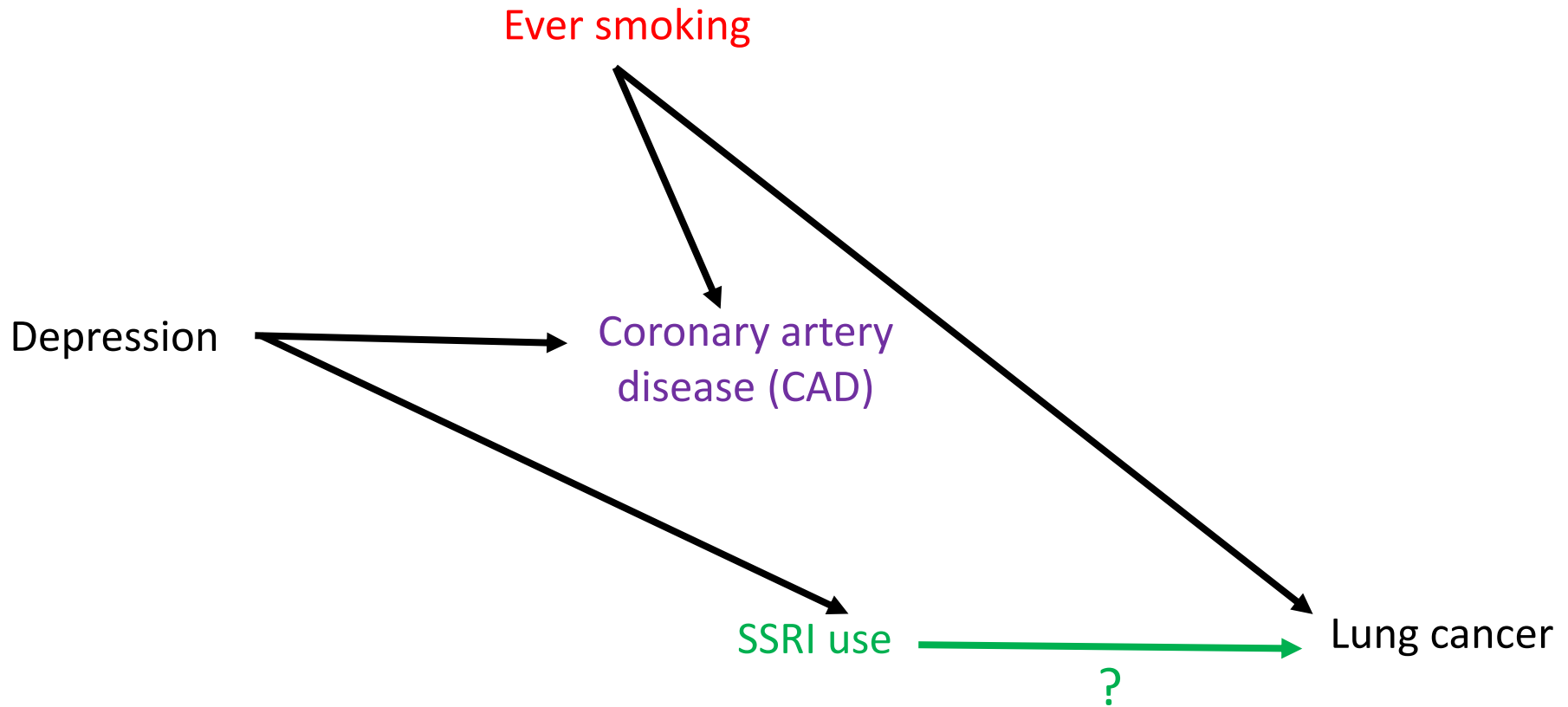
DAGitty online

<https://www.dagitty.net/dags.html#>



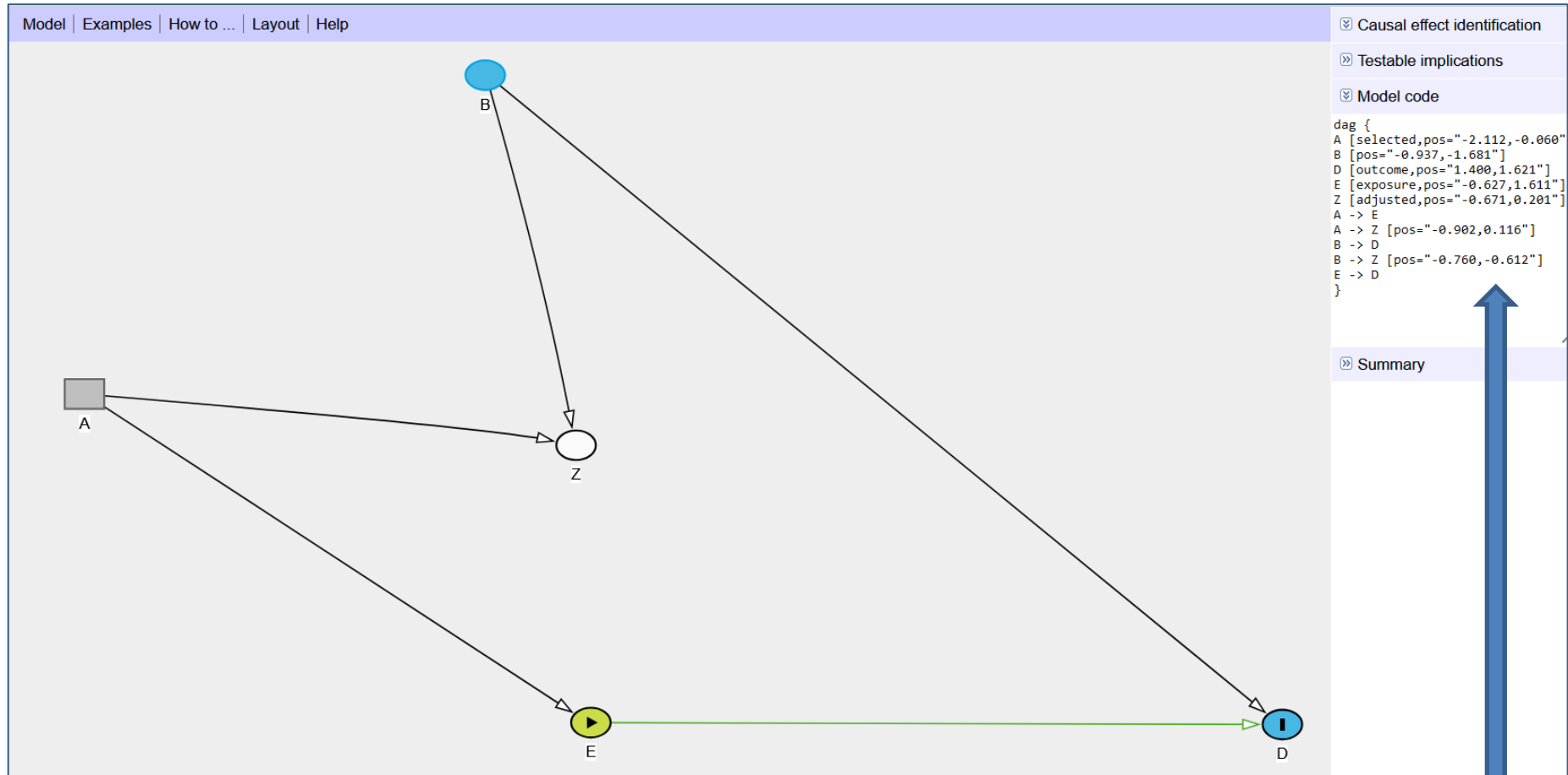
Example 1

Use of selective serotonin reuptake inhibitors (SSRIs) and lung cancer



Implement Example 1 in DAGitty

Move variables so that all arcs flows in the same direction

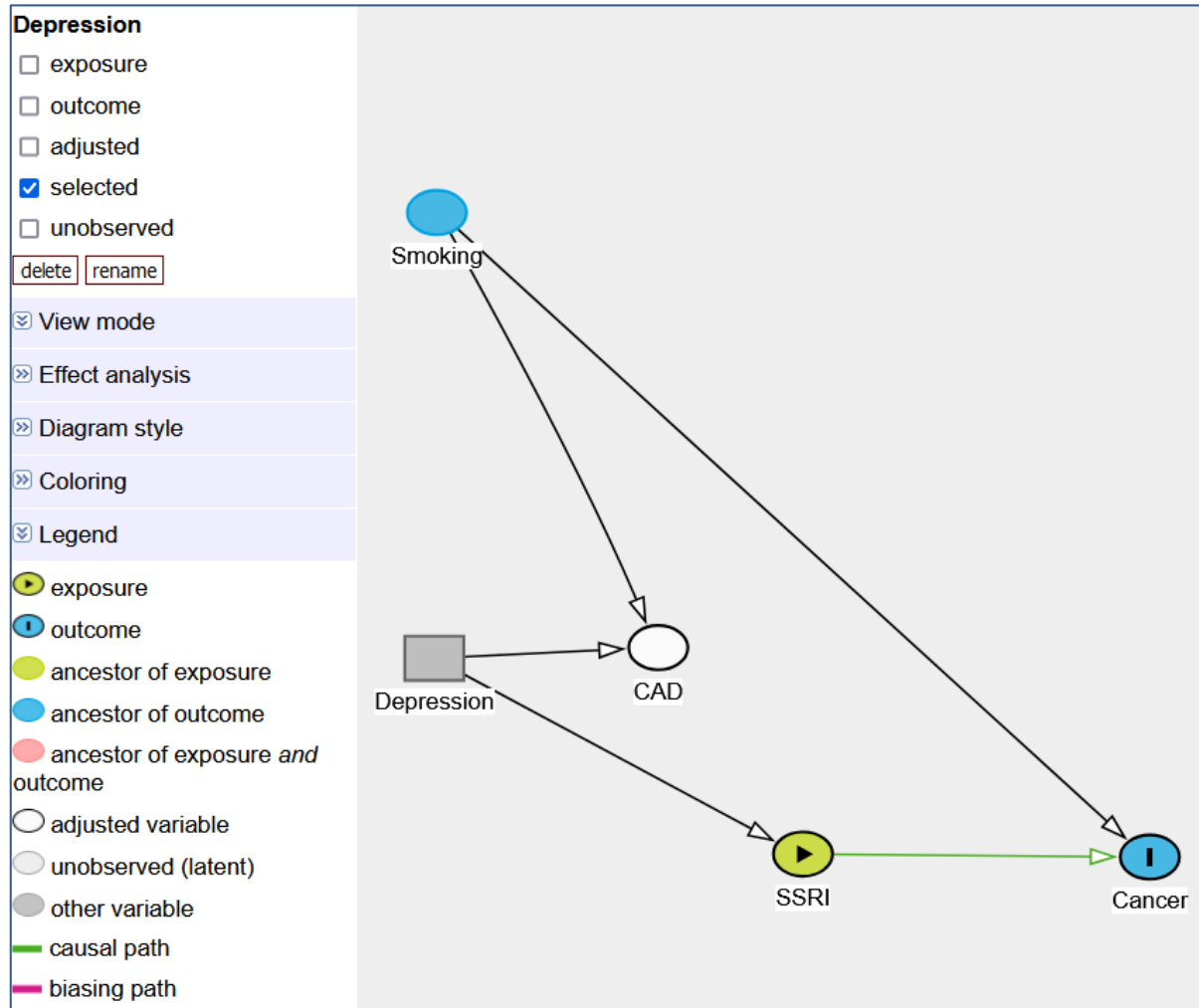


Save model code in a doc!

Model code

```
dag {  
  A [selected, pos="-1.967,0.301"]  
  B [pos="-1.093,-1.621"]  
  D [outcome, pos="1.400,1.621"]  
  E [exposure, pos="-0.286,1.576"]  
  Z [adjusted, pos="-0.202,0.437"]  
  A -> E  
  A -> Z [pos="-0.734,0.377"]  
  B -> D  
  B -> Z [pos="-0.497,-0.321"]  
  E -> D  
}
```

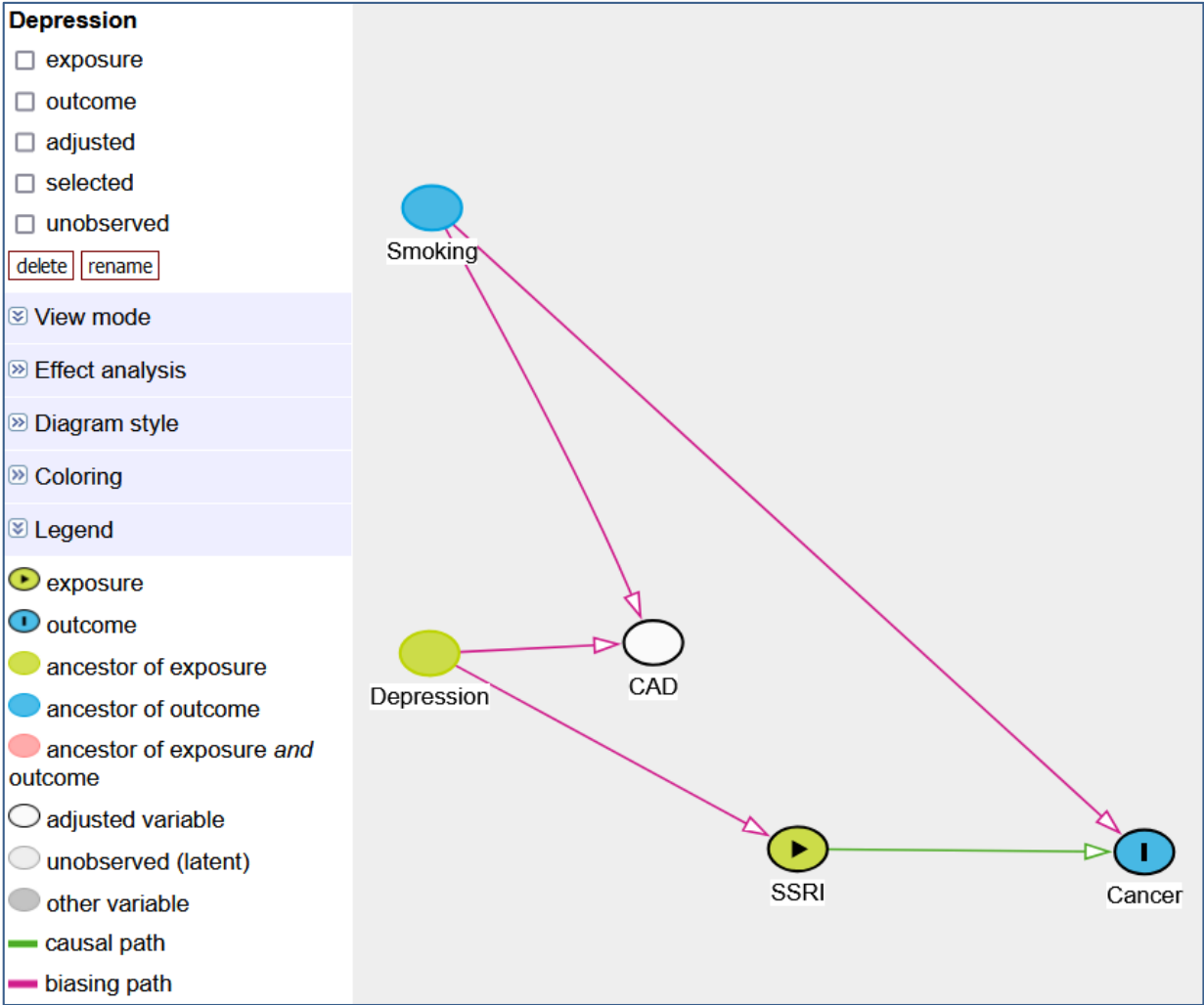
Rename variables



```

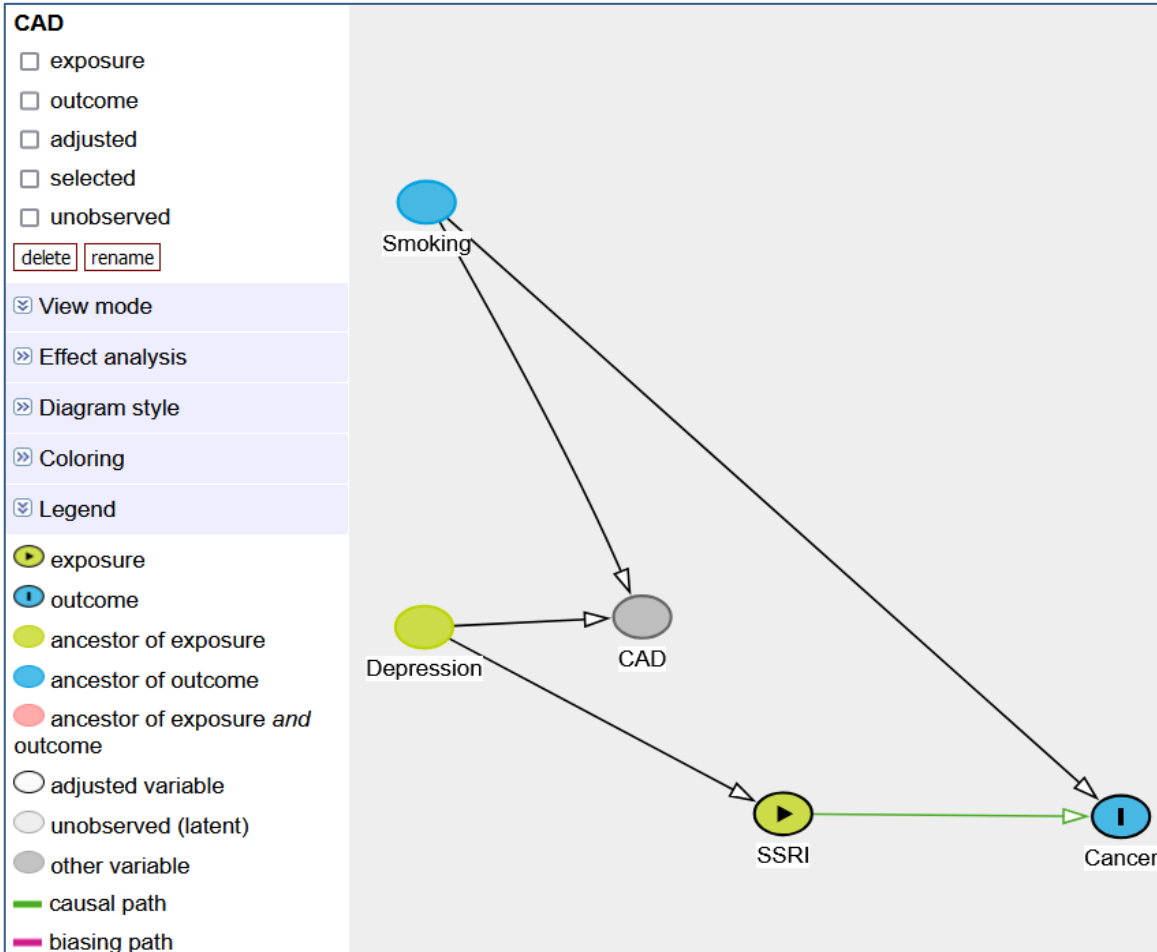
dag {
  CAD [adjusted,pos="-0.995,-0.291"]
  Cancer [outcome,pos="0.172,0.352"]
  Depression [selected,pos="-1.811,-0.446"]
  SSRI [exposure,pos="-0.714,0.352"]
  Smoking [pos="-1.817,-1.656"]
  Depression -> CAD [pos="-1.177,-0.341"]
  Depression -> SSRI
  SSRI -> Cancer
  Smoking -> CAD [pos="-1.154,-0.612"]
  Smoking -> Cancer}
  
```


Unselect Depression



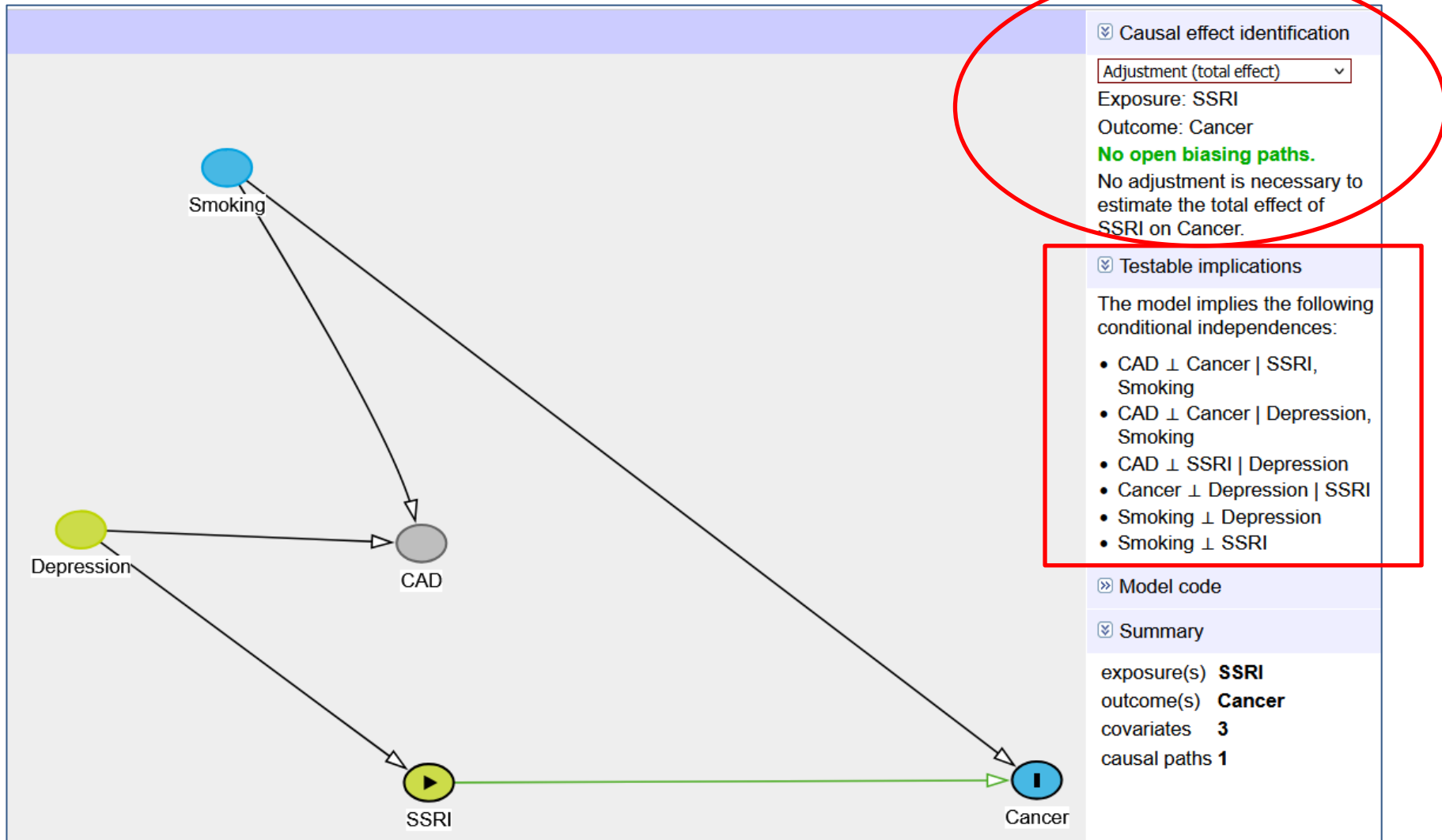
```
dag {
  CAD [adjusted,pos="-0.995,-0.291"]
  Cancer [outcome,pos="0.172,0.352"]
  Depression [pos="-1.811,-0.446"]
  SSRI [exposure,pos="-0.714,0.352"]
  Smoking [pos="-1.817,-1.656"]
  Depression -> CAD [pos="-1.177,-0.341"]
  Depression -> SSRI
  SSRI -> Cancer
  Smoking -> CAD [pos="-1.154,-0.612"]
  Smoking -> Cancer}
```

Unadjust for CAD

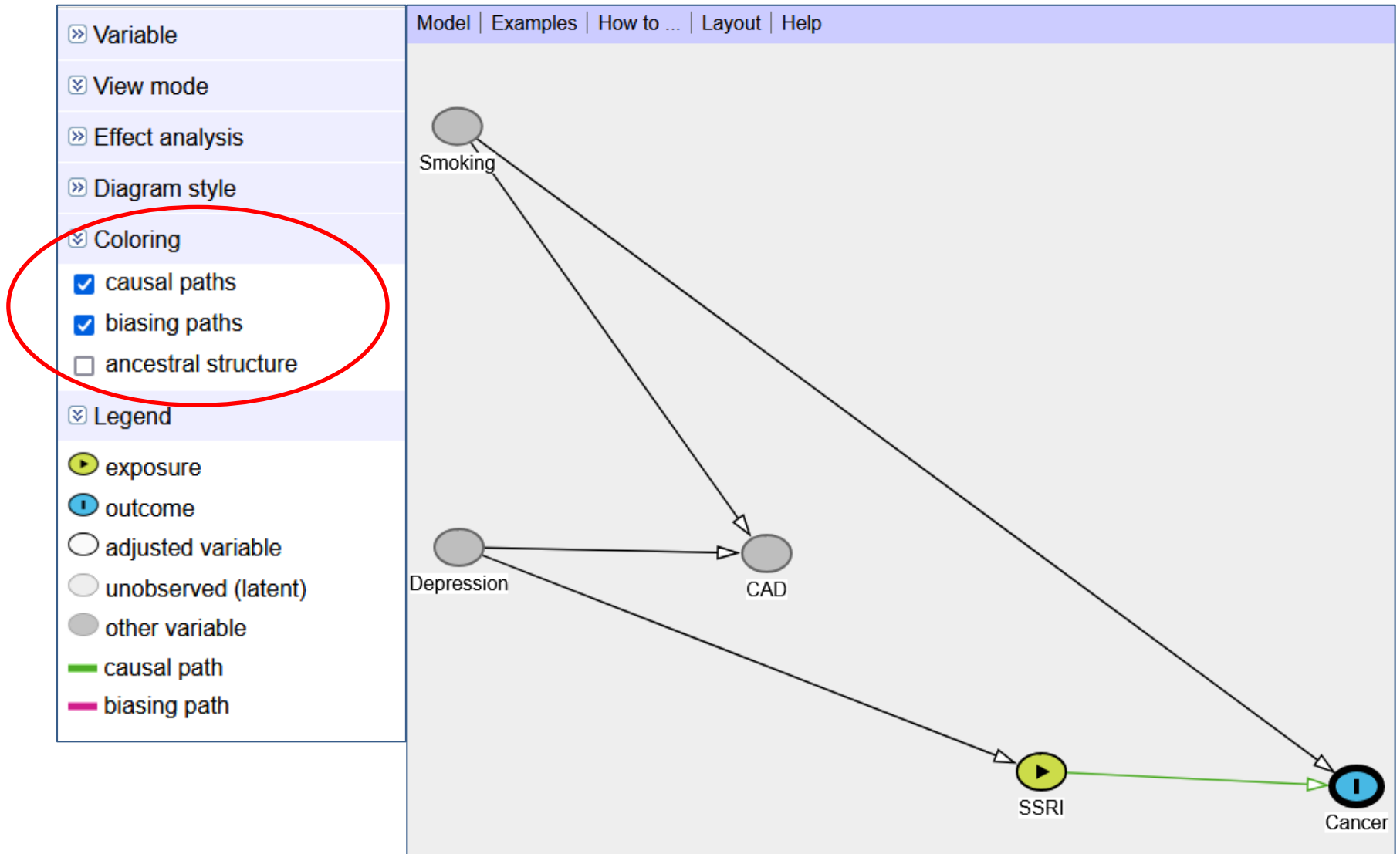


```
dag {  
  CAD [pos="-0.995,-0.291"]  
  Cancer [outcome,pos="0.172,0.352"]  
  Depression [pos="-1.811,-0.446"]  
  SSRI [exposure,pos="-0.714,0.352"]  
  Smoking [pos="-1.817,-1.656"]  
  Depression -> CAD  
  Depression -> SSRI  
  SSRI -> Cancer  
  Smoking -> CAD [pos="-1.154,-0.612"]  
  Smoking -> Cancer  
}
```

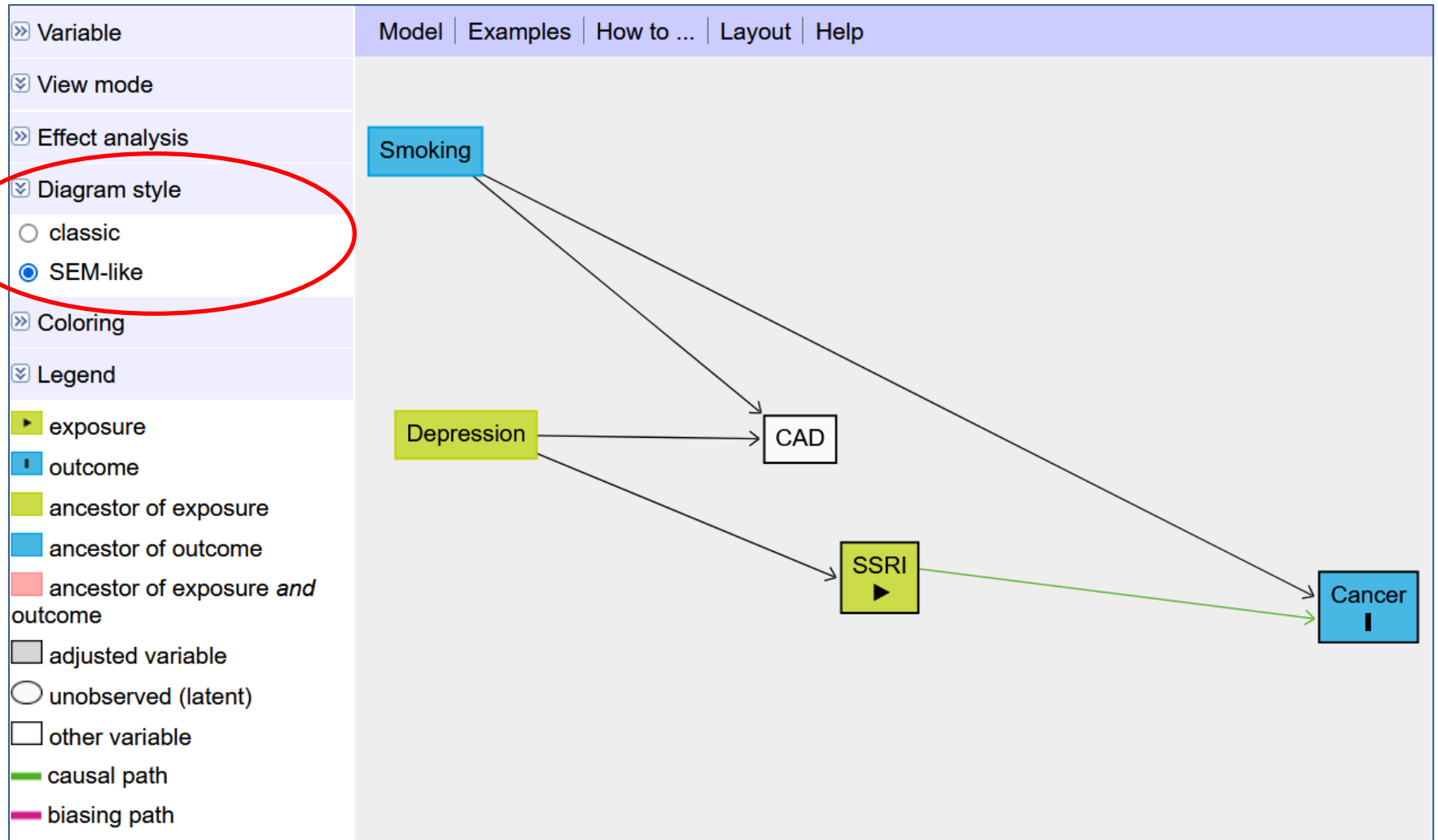
Look at the results



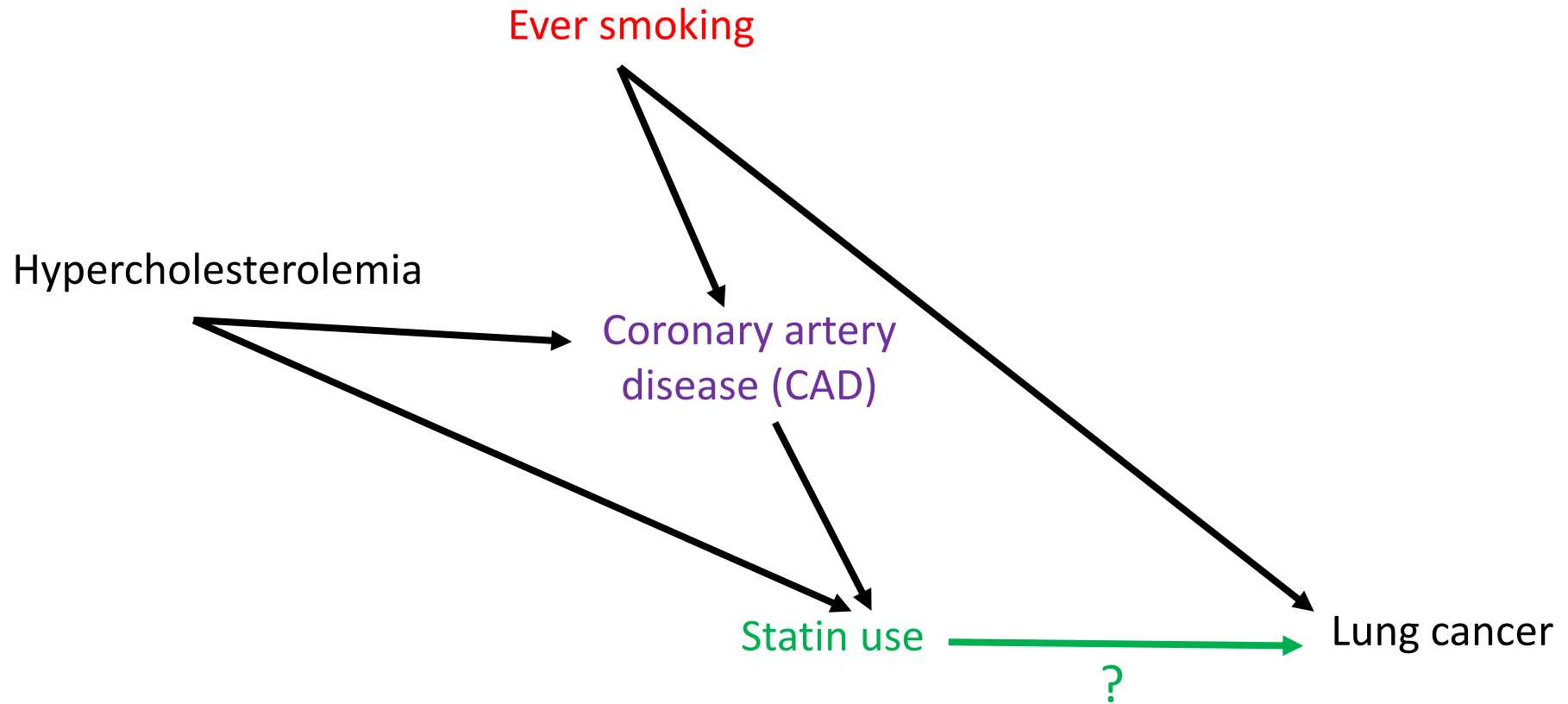
Possibility to not color ancestral structure



Possibility to display the DAG like a SEM

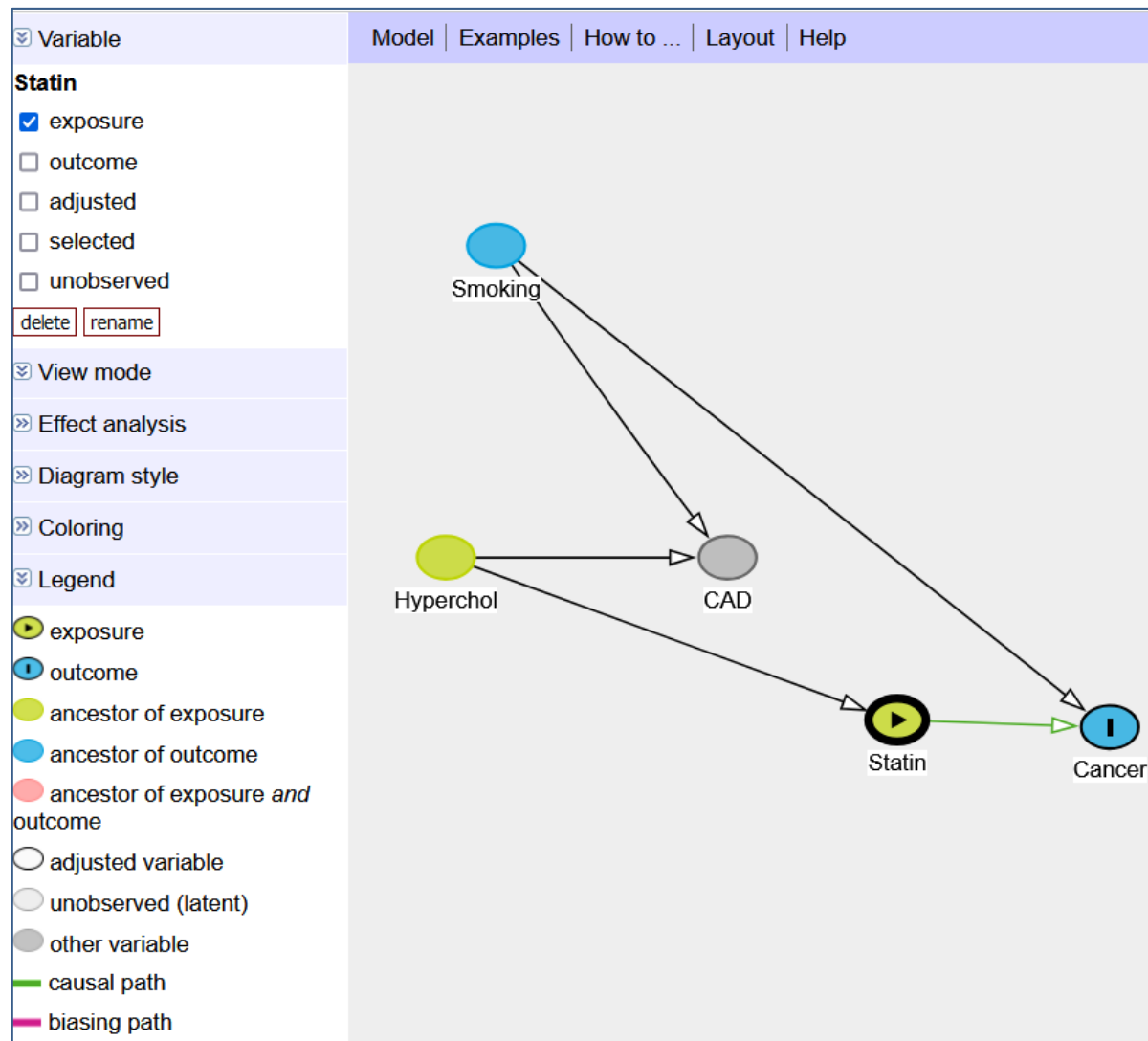


Example 2: Use of statin and lung cancer



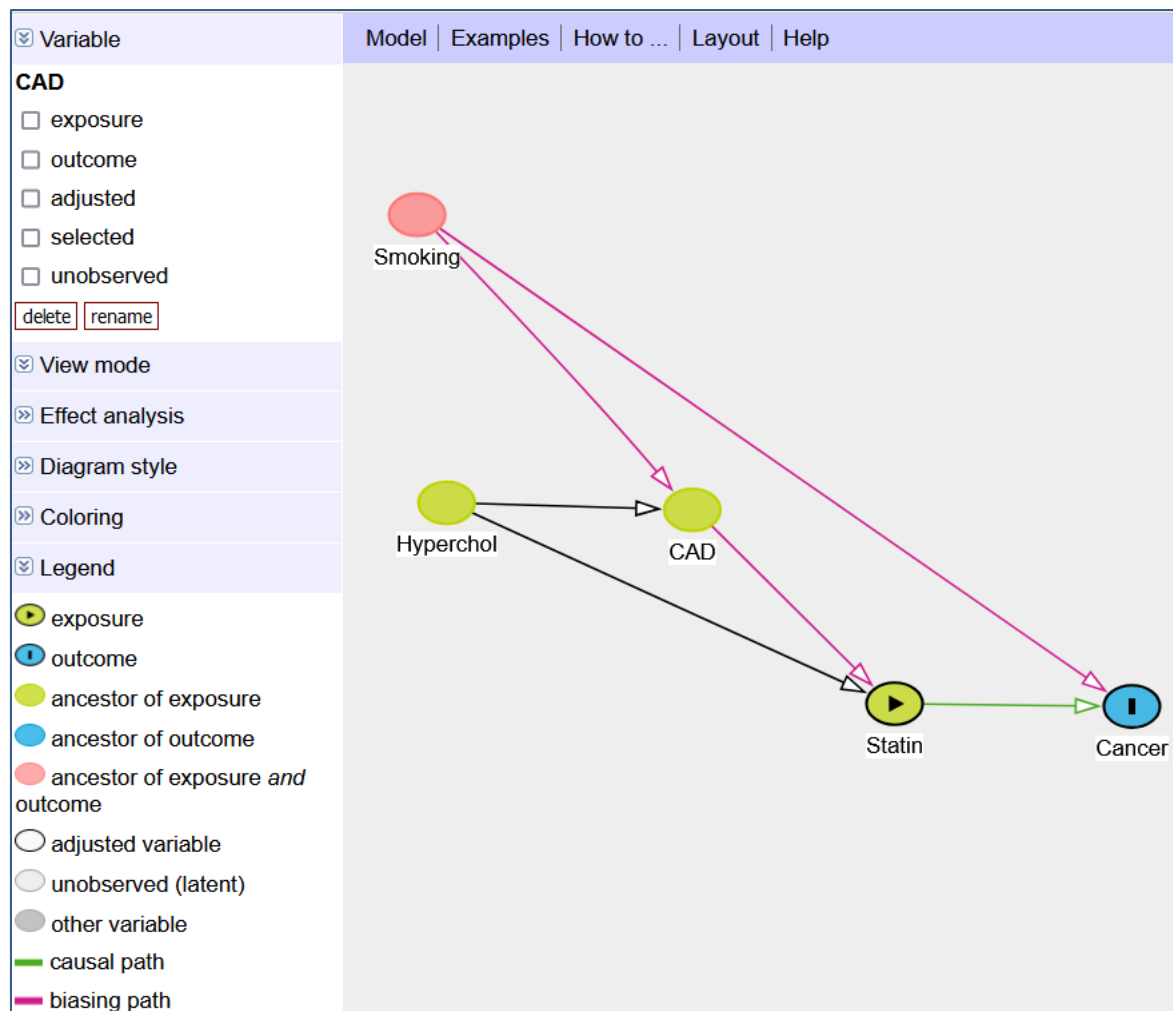
Take the code for the DAG of Example 1

Rename « Depression » as « Hyperchol » and « SSRI » as « Statin »



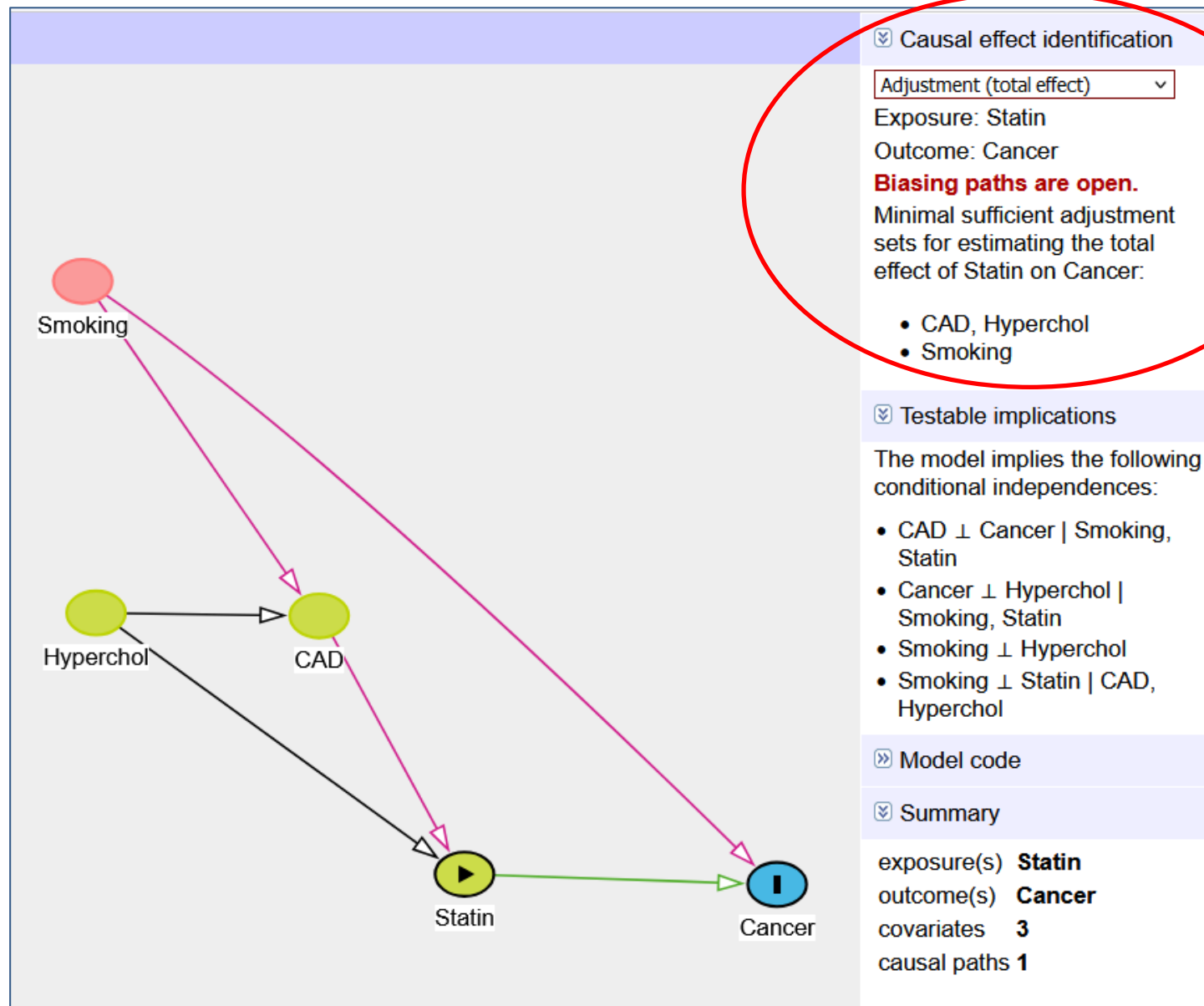
```
dag {  
  CAD [pos="-1.568,-0.469"]  
  Cancer [outcome,pos="-1.036,-0.014"]  
  Hyperchol [pos="-1.811,-0.446"]  
  Smoking [pos="-1.808,-1.416"]  
  Statin [exposure,pos="-1.412,-0.020"]  
  Hyperchol -> CAD  
  Hyperchol -> Statin  
  Smoking -> CAD [pos="-1.642,-0.822"]  
  Smoking -> Cancer  
  Statin -> Cancer  
}
```

Add an arrow between CAD and statin



```
dag {  
  CAD [pos="-1.568,-0.469"]  
  Cancer [outcome,pos="-1.036,-0.014"]  
  Hyperchol [pos="-1.811,-0.446"]  
  Smoking [pos="-1.808,-1.416"]  
  Statin [exposure,pos="-1.412,-0.020"]  
  CAD -> Statin  
  Hyperchol -> CAD  
  Hyperchol -> Statin  
  Smoking -> CAD [pos="-1.642,-0.822"]  
  Smoking -> Cancer  
  Statin -> Cancer}
```

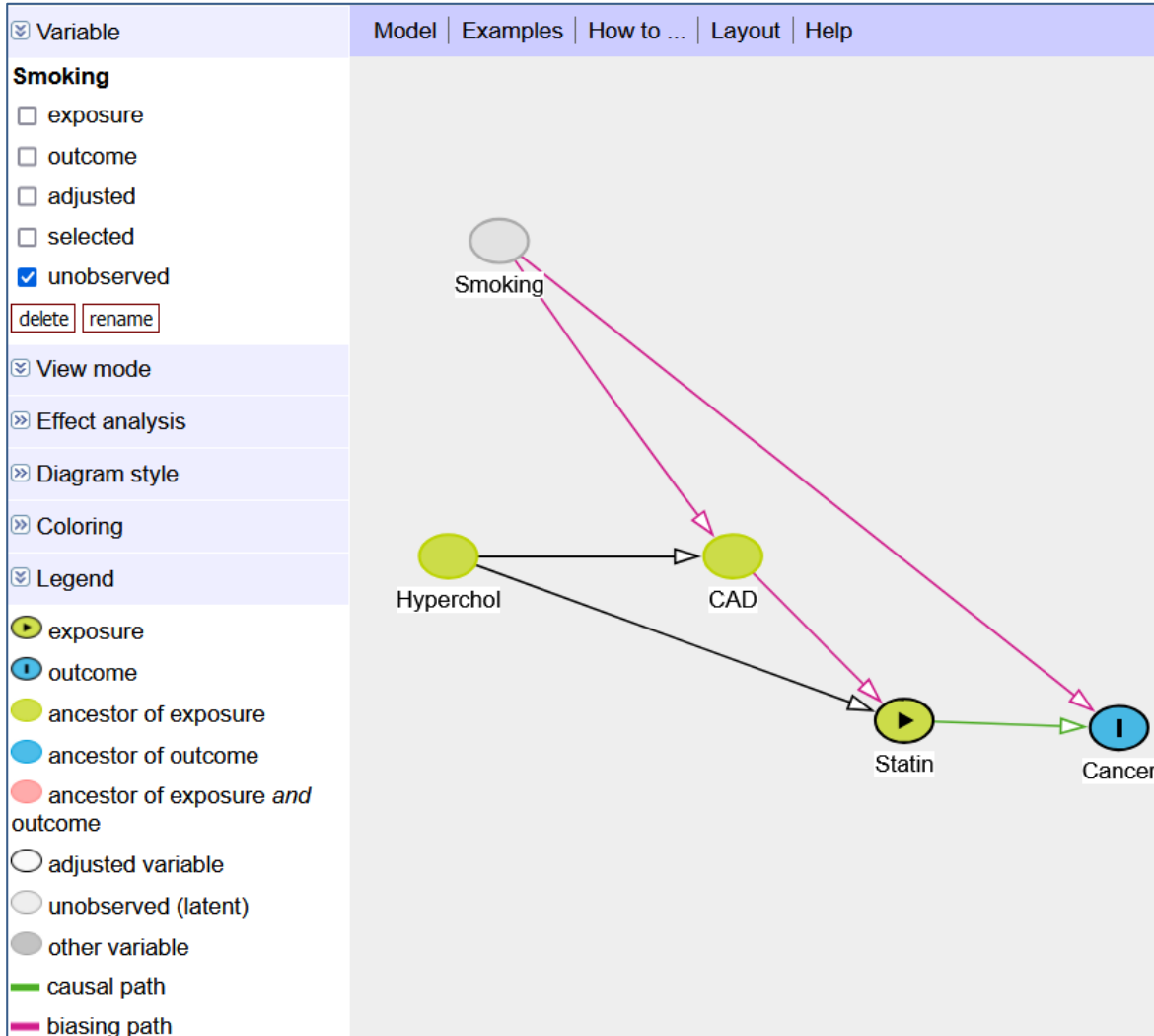

Look at the results



Unobserved (latent) variables

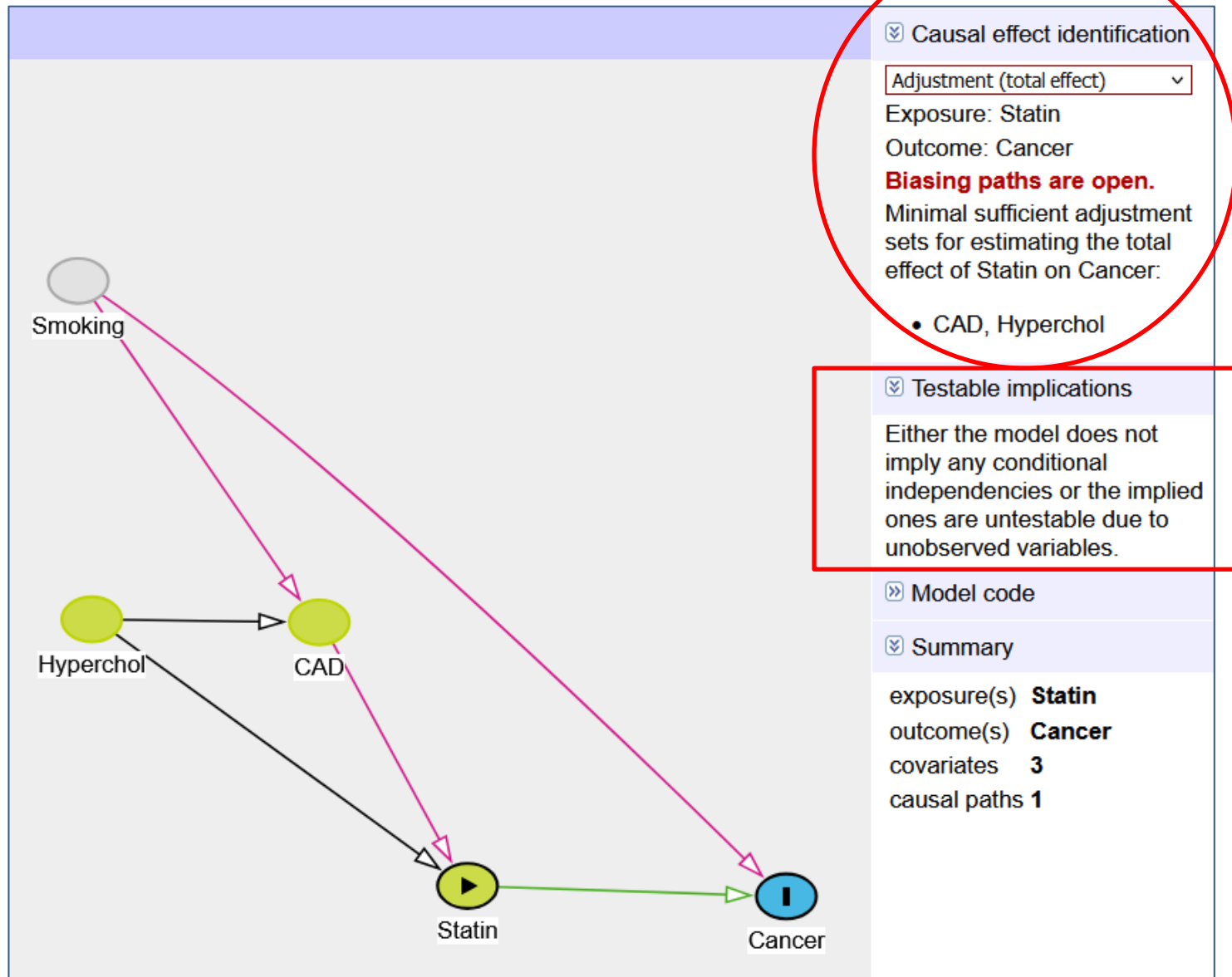
Take Example 2

Indicate that smoking is not observed



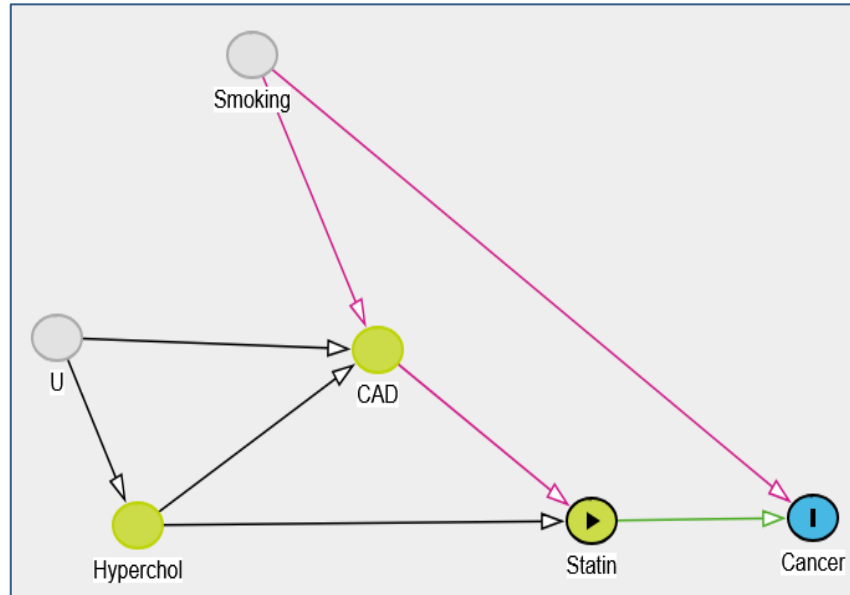
```
dag {  
  CAD [pos="-1.472,-0.764"]  
  Cancer [outcome,pos="-1.047,-0.386"]  
  Hyperchol [pos="-1.786,-0.764"]  
  Smoking [latent,pos="-1.730,-1.460"]  
  Statin [exposure,pos="-1.284,-0.402"]  
  CAD -> Statin  
  Hyperchol -> CAD  
  Hyperchol -> Statin  
  Smoking -> CAD [pos="-1.597,-1.082"]  
  Smoking -> Cancer  
  Statin -> Cancer  
}
```

Look at the results

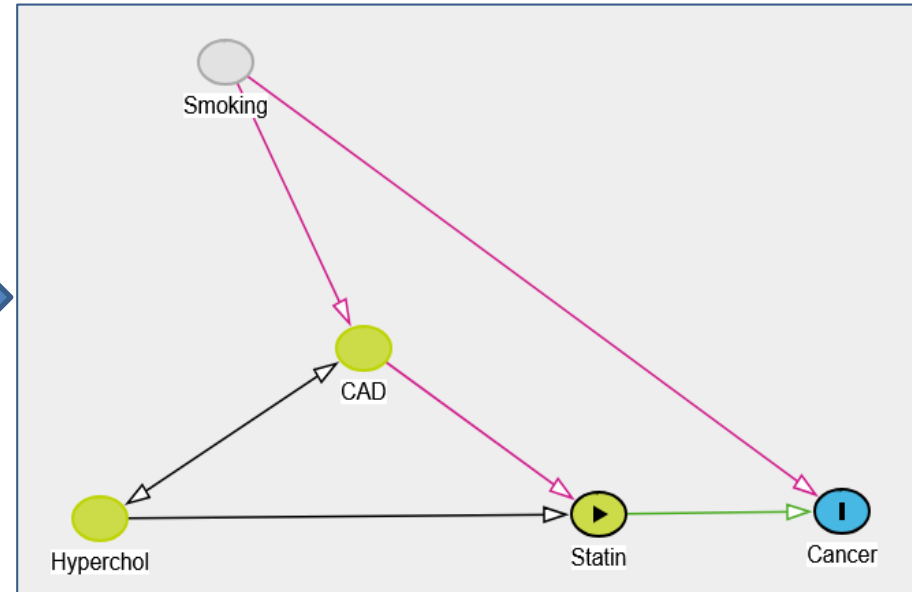


If several unknown confounders between two variables

Example: between hyperchol and CAD

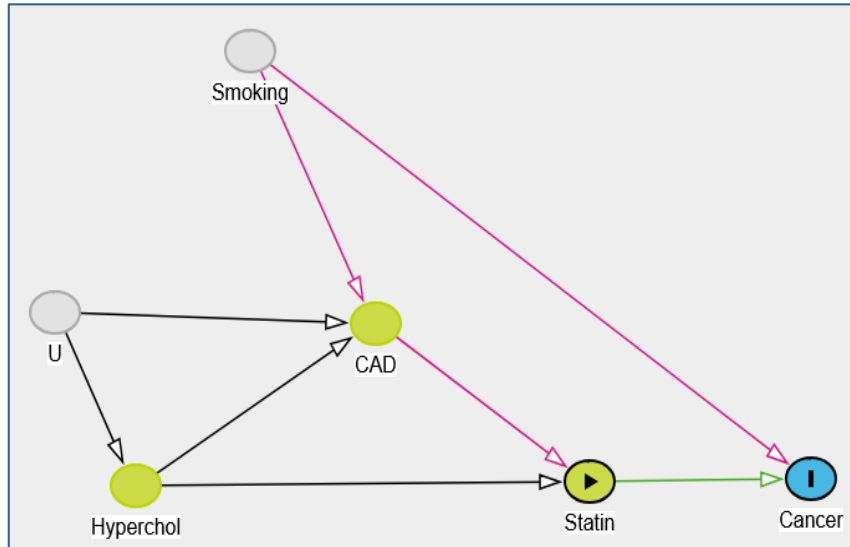


```
dag {
  CAD [pos="-1.398,-0.977"]
  Cancer [outcome,pos="-1.205,-0.817"]
  Hyperchol [pos="-1.504,-0.810"]
  Smoking [latent,pos="-1.463,-1.254"]
  Statin [exposure,pos="-1.303,-0.814"]
  U [latent,pos="-1.479,-0.977"]
  CAD -> Statin
  Hyperchol -> CAD
  Hyperchol -> Statin
  Smoking -> CAD
  Smoking -> Cancer
  Statin -> Cancer
  U -> CAD
  U -> Hyperchol}
```



```
dag {
  CAD [pos="-1.398,-0.977"]
  Cancer [outcome,pos="-1.205,-0.817"]
  Hyperchol [pos="-1.504,-0.810"]
  Smoking [latent,pos="-1.454,-1.258"]
  Statin [exposure,pos="-1.303,-0.814"]
  CAD -> Statin
  CAD <-> Hyperchol
  Hyperchol -> Statin
  Smoking -> CAD
  Smoking -> Cancer
  Statin -> Cancer}
```

Look at the results



☑ Causal effect identification

Adjustment (total effect) ▾

Exposure: Statin

Outcome: Cancer

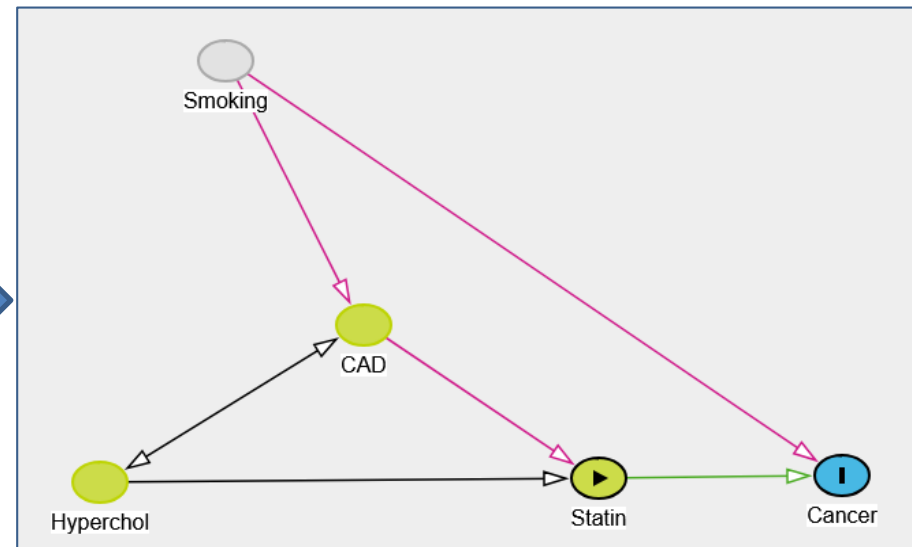
Biasing paths are open.

Minimal sufficient adjustment sets for estimating the total effect of Statin on Cancer:

- CAD, Hyperchol

☑ Testable implications

Either the model does not imply any conditional independencies or the implied ones are untestable due to unobserved variables.



☑ Causal effect identification

Adjustment (total effect) ▾

Exposure: Statin

Outcome: Cancer

Biasing paths are open.

Minimal sufficient adjustment sets for estimating the total effect of Statin on Cancer:

- CAD, Hyperchol

☑ Testable implications

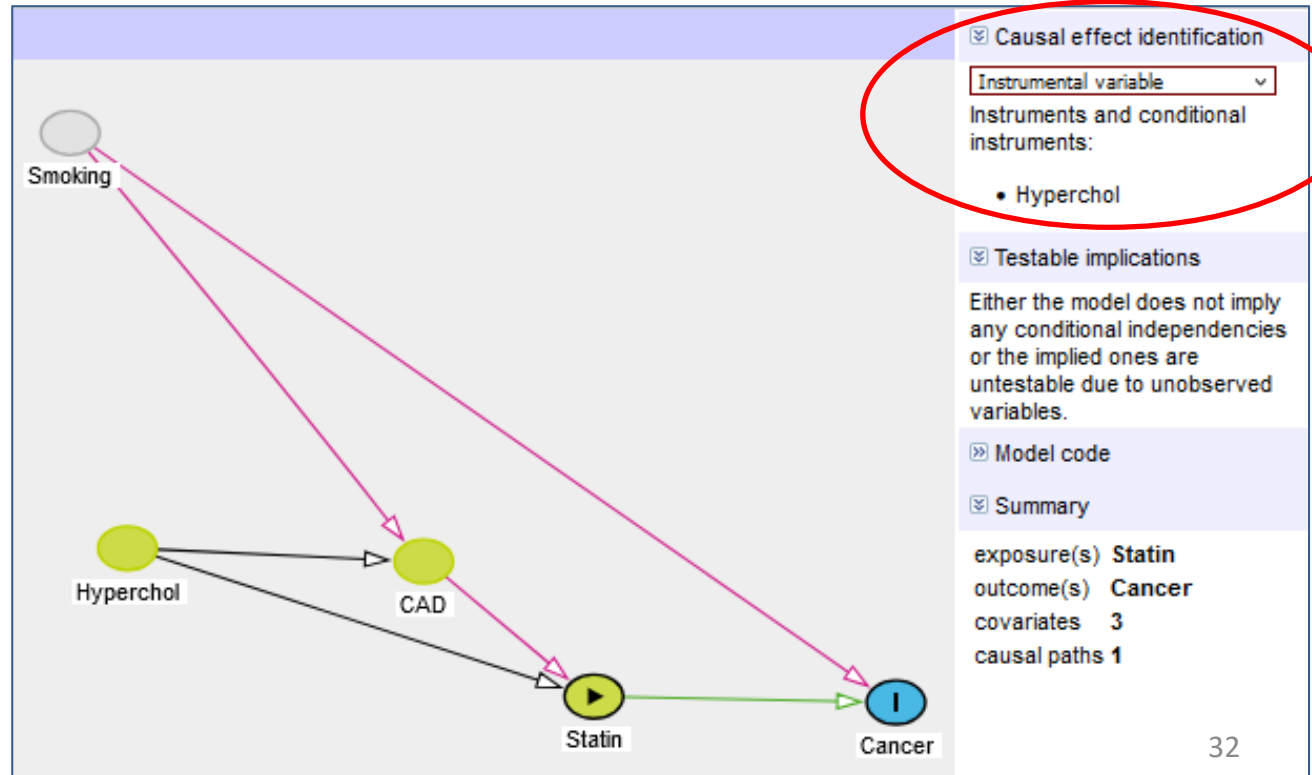
Either the model does not imply any conditional independencies or the implied ones are untestable due to unobserved variables.

Instrumental variables

Ask to display instrumental variables

Take Example 2: Smoking still unobserved

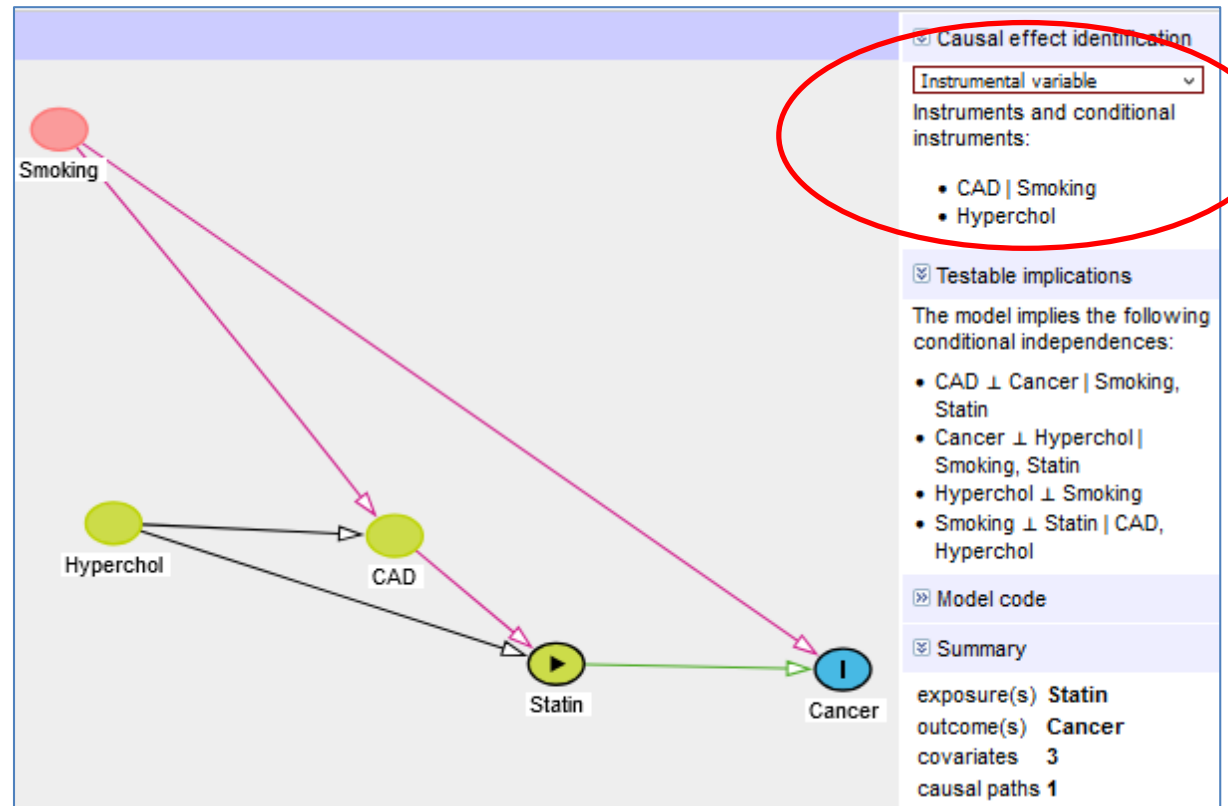
- Instrumental variable (or instrument I)
 - There must be an open path between I and the exposure X (exogeneity); and
 - All paths between I and the outcome Y must be closed in a modified graph where all edges out of X are removed (exclusion restriction)
- Conditional instrument
 - There must be a path between I and X that is opened by Z ; and
 - All paths between I and Y must be closed by Z in a modified graph where all edges out of X are removed.



Ask to display instrumental variables

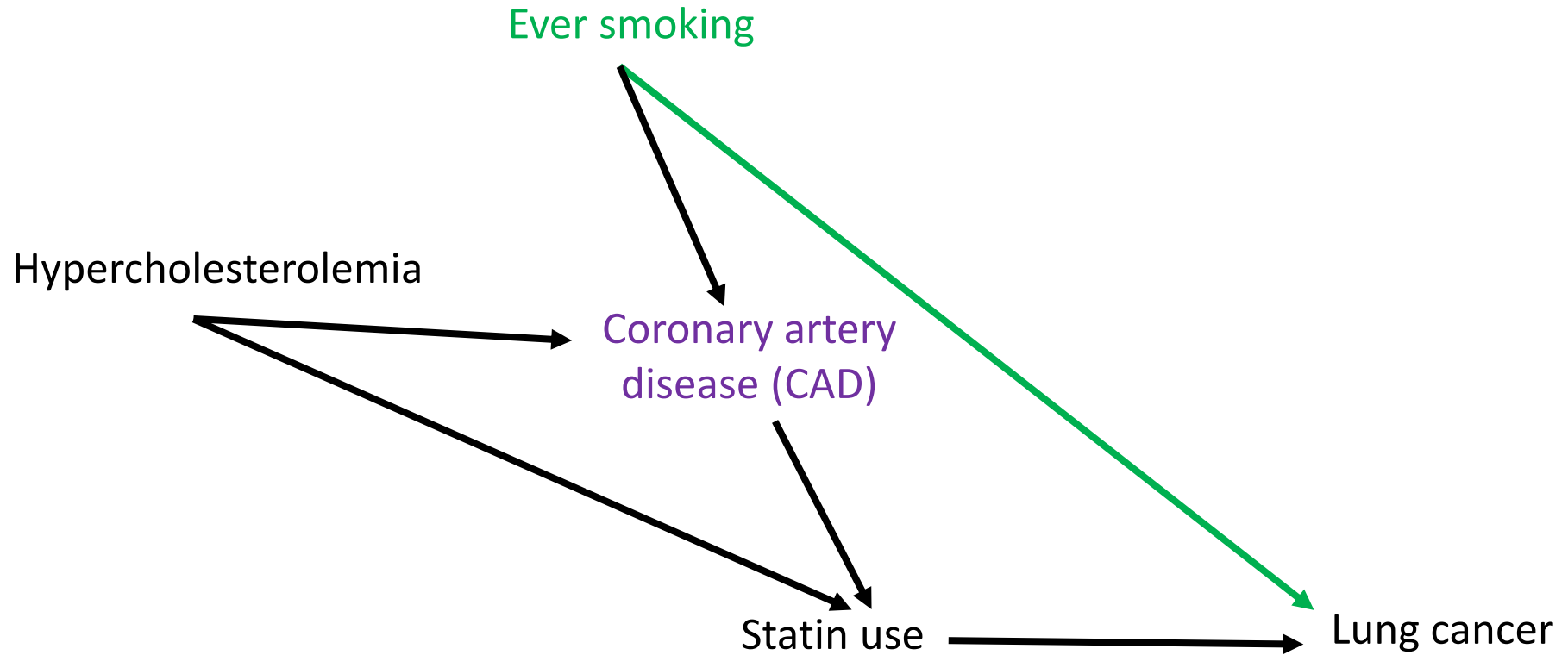
Take Example 2: Smoking observed

- Instrumental variable (or instrument I)
 1. There must be an open path between I and the exposure X (exogeneity); and
 2. All paths between I and the outcome Y must be closed in a modified graph where all edges out of X are removed (exclusion restriction)
- Conditional instrument
 1. There must be a path between I and X that is opened by Z ; and
 2. All paths between I and Y must be closed by Z in a modified graph where all edges out of X are removed.



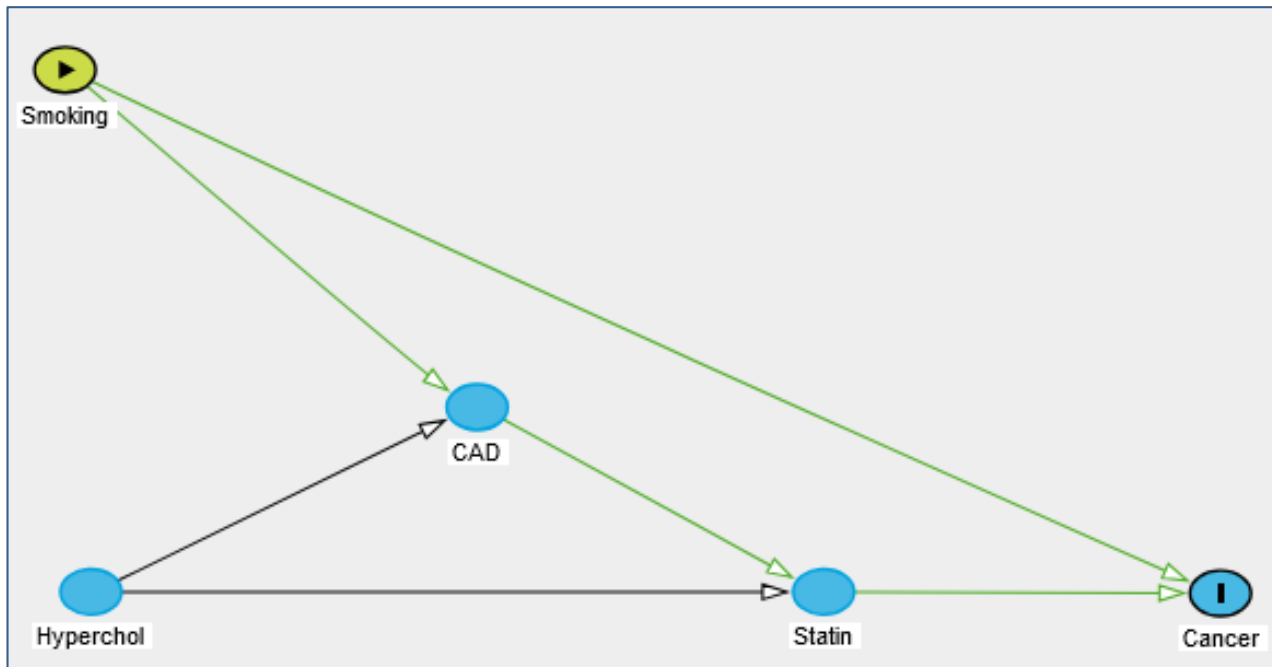
Direct effects

Example 3: Smoking and lung cancer



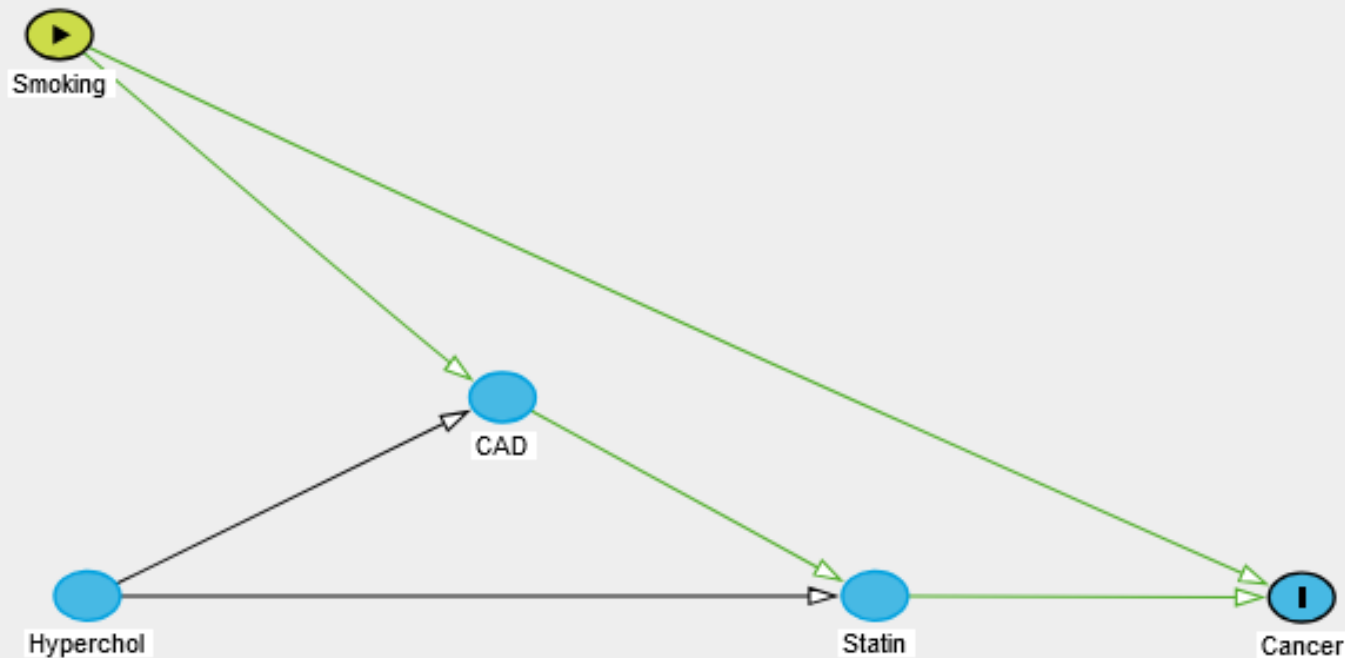
Take the model code for DAG of Example 2

Set Smoking as the observed Exposure



```
dag {  
  CAD [pos="-1.287,-1.077"]  
  Cancer [outcome,pos="-1.050,-0.923"]  
  Hyperchol [pos="-1.391,-0.902"]  
  Smoking [exposure,pos="-1.394,-1.420"]  
  Statin [pos="-1.182,-0.930"]  
  CAD -> Statin  
  Hyperchol -> CAD  
  Hyperchol -> Statin  
  Smoking -> CAD [pos="-1.348,-1.268"]  
  Smoking -> Cancer [pos="-1.235,-1.204"]  
  Statin -> Cancer}
```

Look at the results



☑ Causal effect identification

Adjustment (total effect) ▾

Exposure: Smoking

Outcome: Cancer

No open biasing paths.

No adjustment is necessary to estimate the total effect of Smoking on Cancer.

☑ Testable implications

The model implies the following conditional independences:

- $CAD \perp\!\!\!\perp Cancer \mid Smoking, Statin$
- $Cancer \perp\!\!\!\perp Hyperchol \mid Smoking, Statin$
- $Hyperchol \perp\!\!\!\perp Smoking$
- $Smoking \perp\!\!\!\perp Statin \mid CAD, Hyperchol$

» Model code

☑ Summary

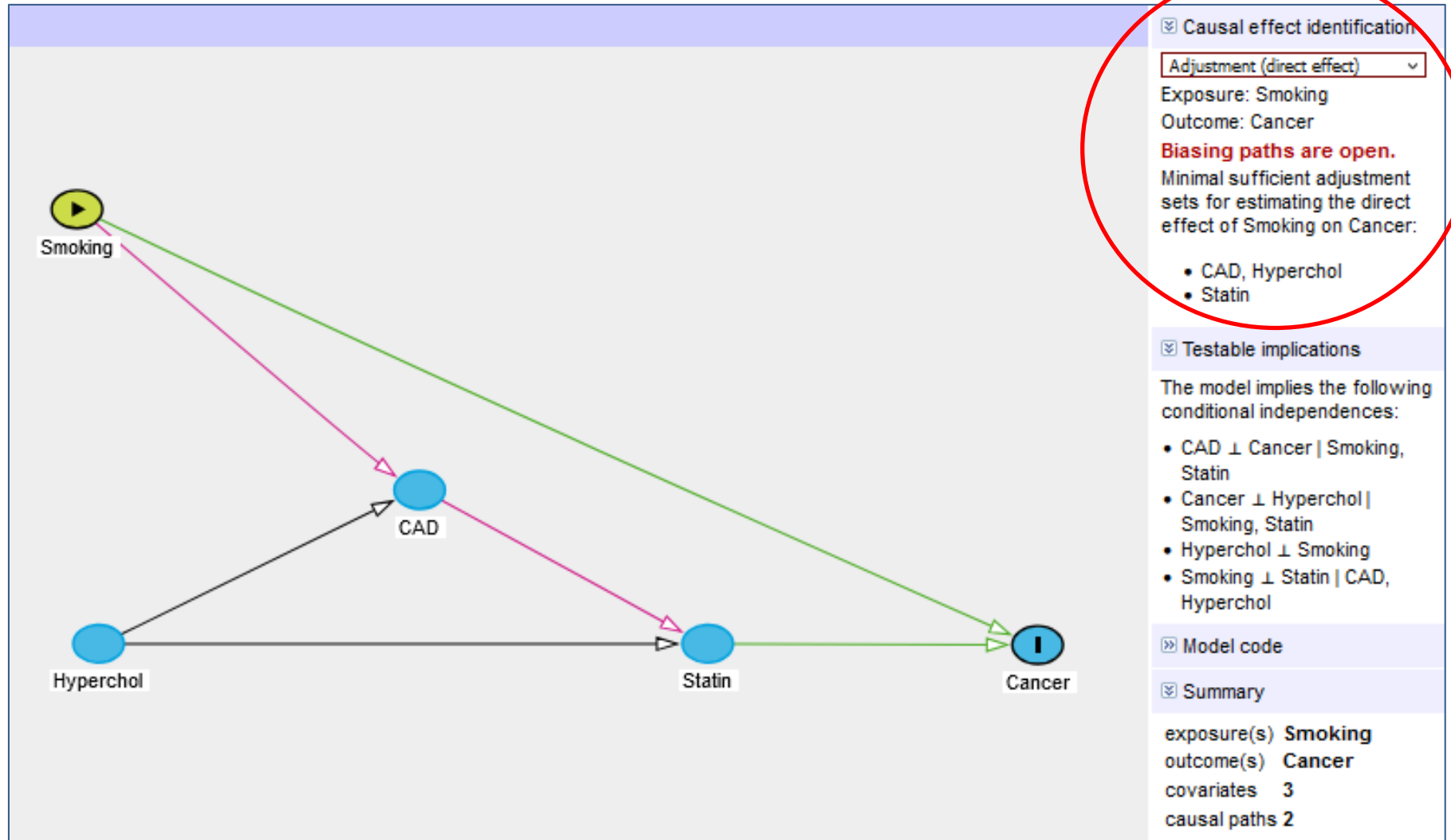
exposure(s) **Smoking**

outcome(s) **Cancer**

covariates 3

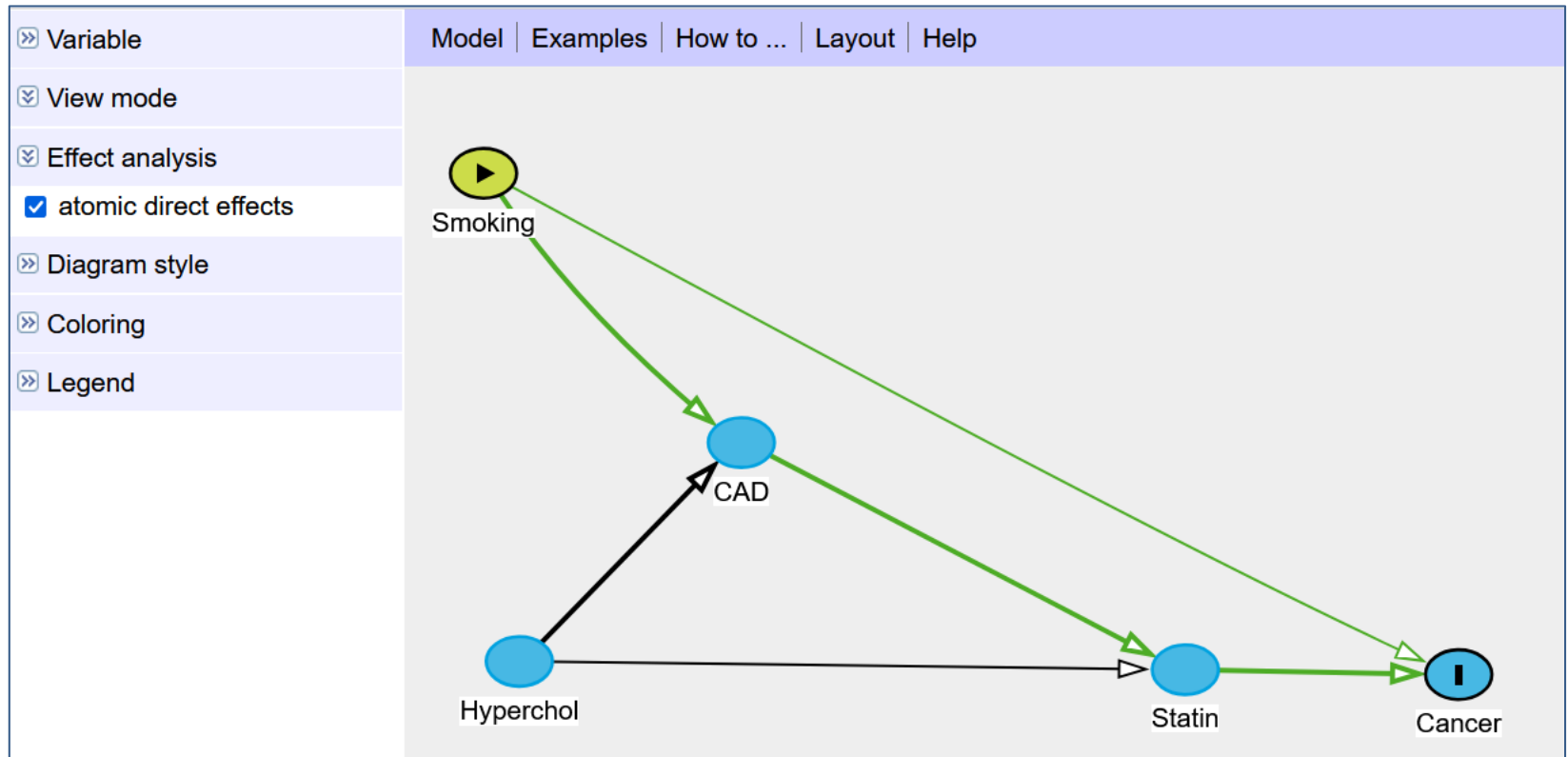
causal paths 2

Display adjustment set to estimate the direct effect of smoking



What are atomic direct effects?

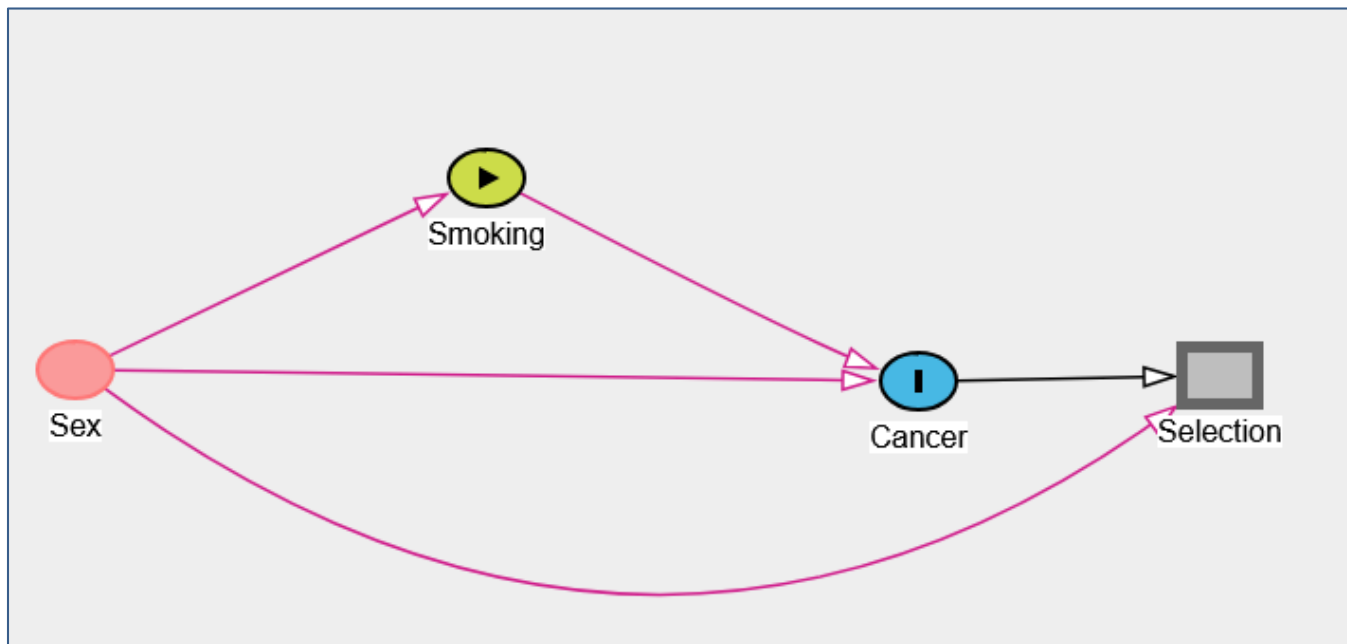
- Arrows (in bold) for which there is no corresponding indirect path
- Removing one of these arrows from the diagram means that there will no longer be any causal effect between the corresponding variables.



Selection bias

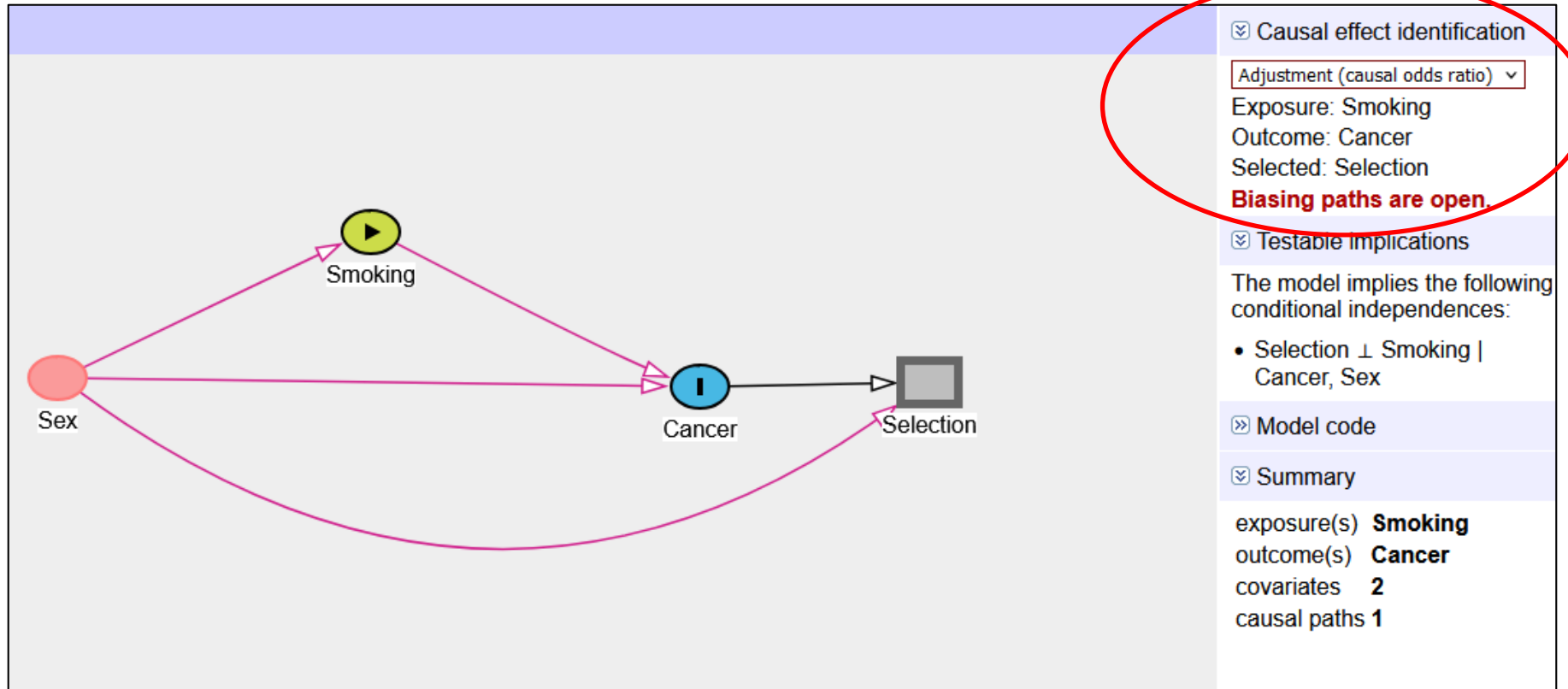
Example 4

A case-control study on smoking and lung cancer, matched on sex, with sex a true confounder

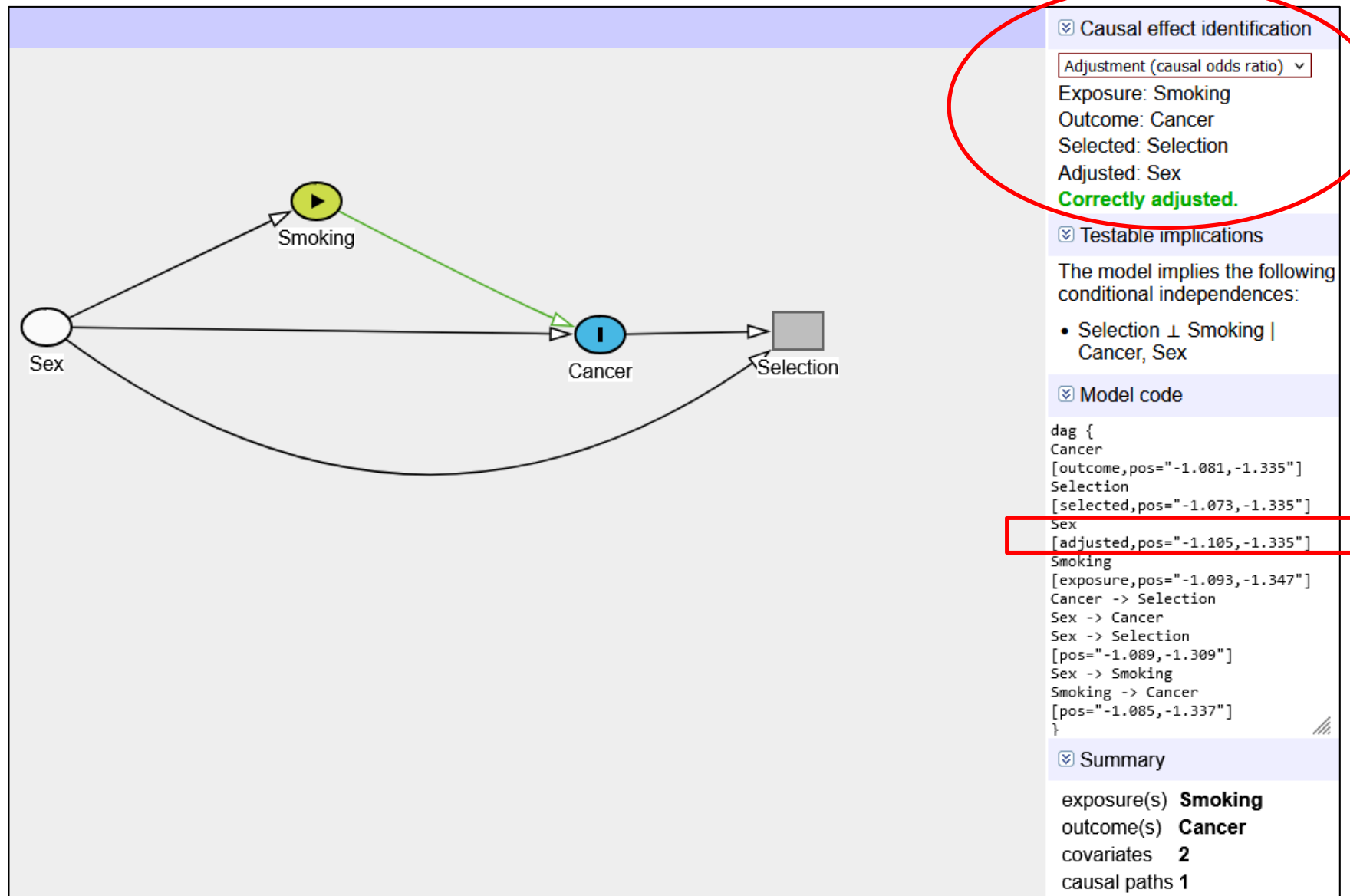


```
dag {  
  Cancer [outcome,pos="-1.081,-1.335"]  
  Selection [selected,pos="-1.073,-1.335"]  
  Sex [pos="-1.105,-1.335"]  
  Smoking [exposure,pos="-1.093,-1.347"]  
  Cancer -> Selection  
  Sex -> Cancer  
  Sex -> Selection [pos="-1.089,-1.309"]  
  Sex -> Smoking  
  Smoking -> Cancer [pos="-1.085,-1.337"]}
```

Look at the results: adjustment for causal odds ratio

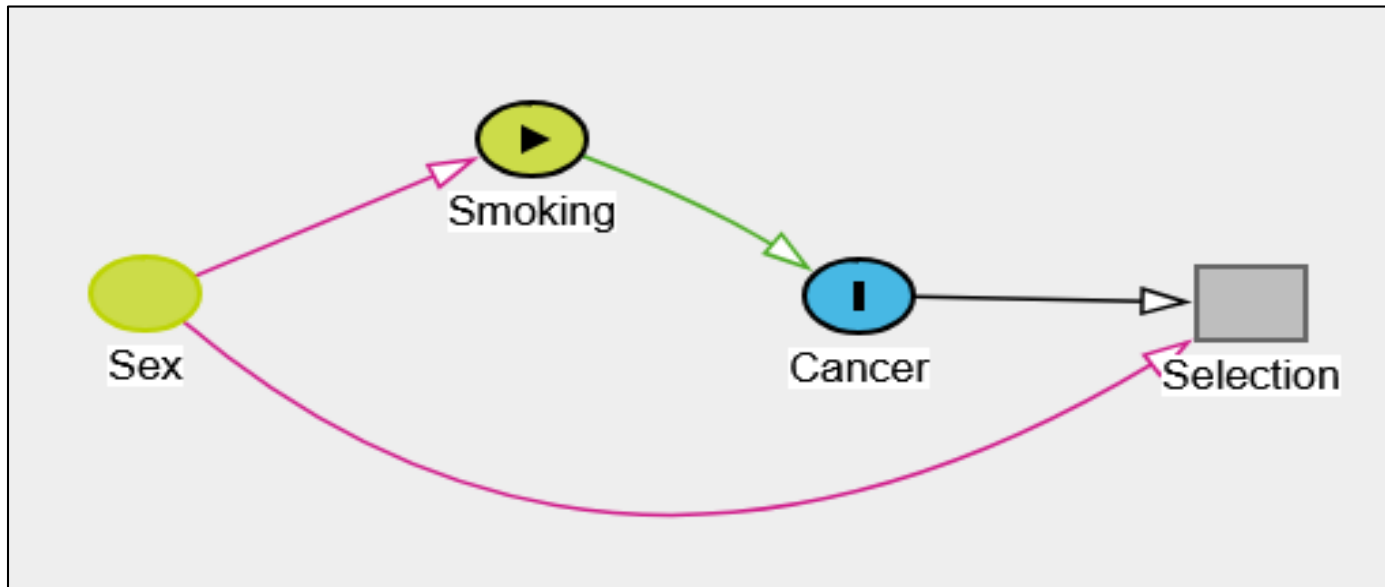


Need to adjust for the matching factor (sex)



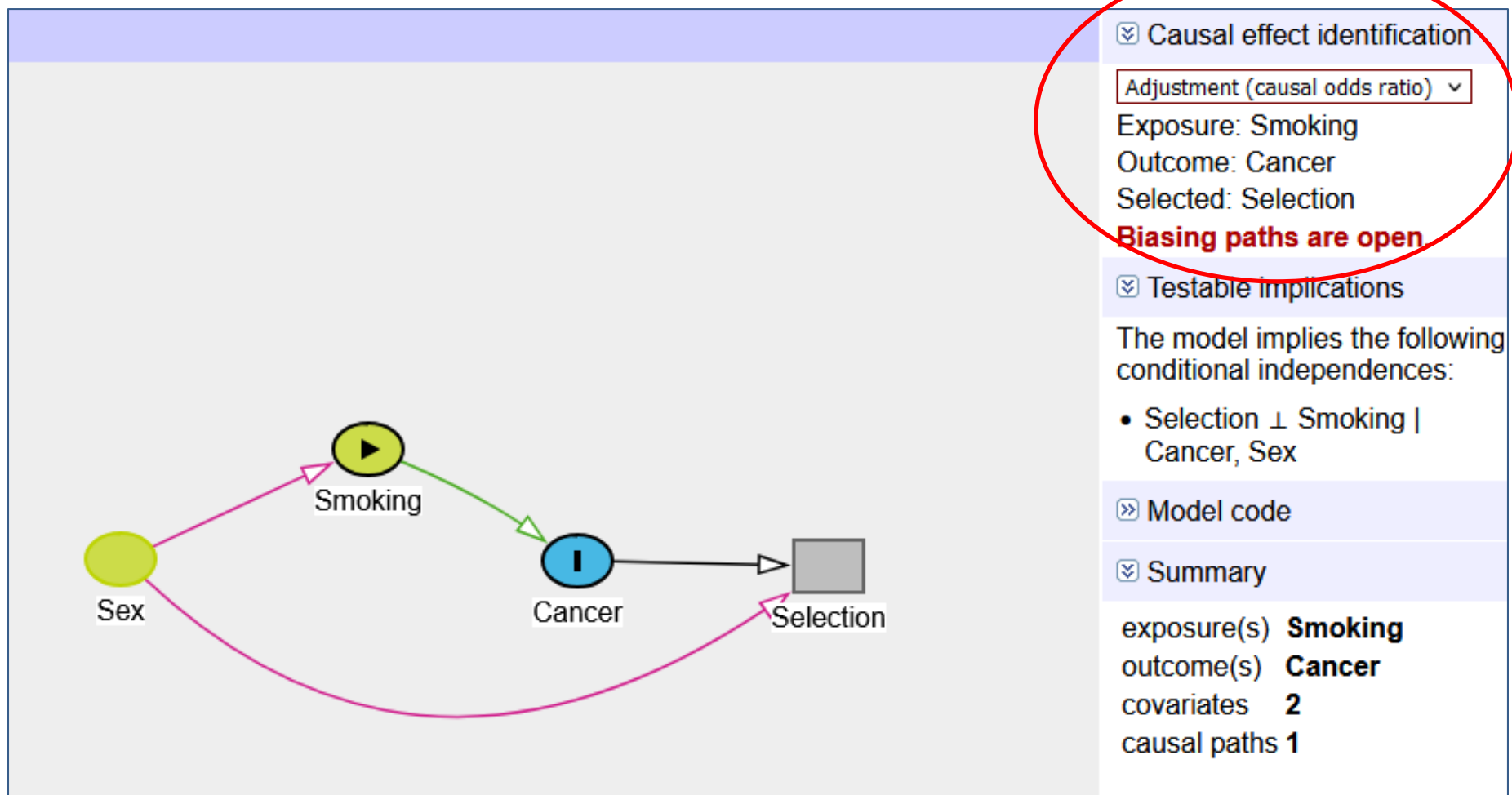
Example 5

A case-control study on smoking and lung cancer, matched on sex, with sex a cause of exposure but not of the outcome

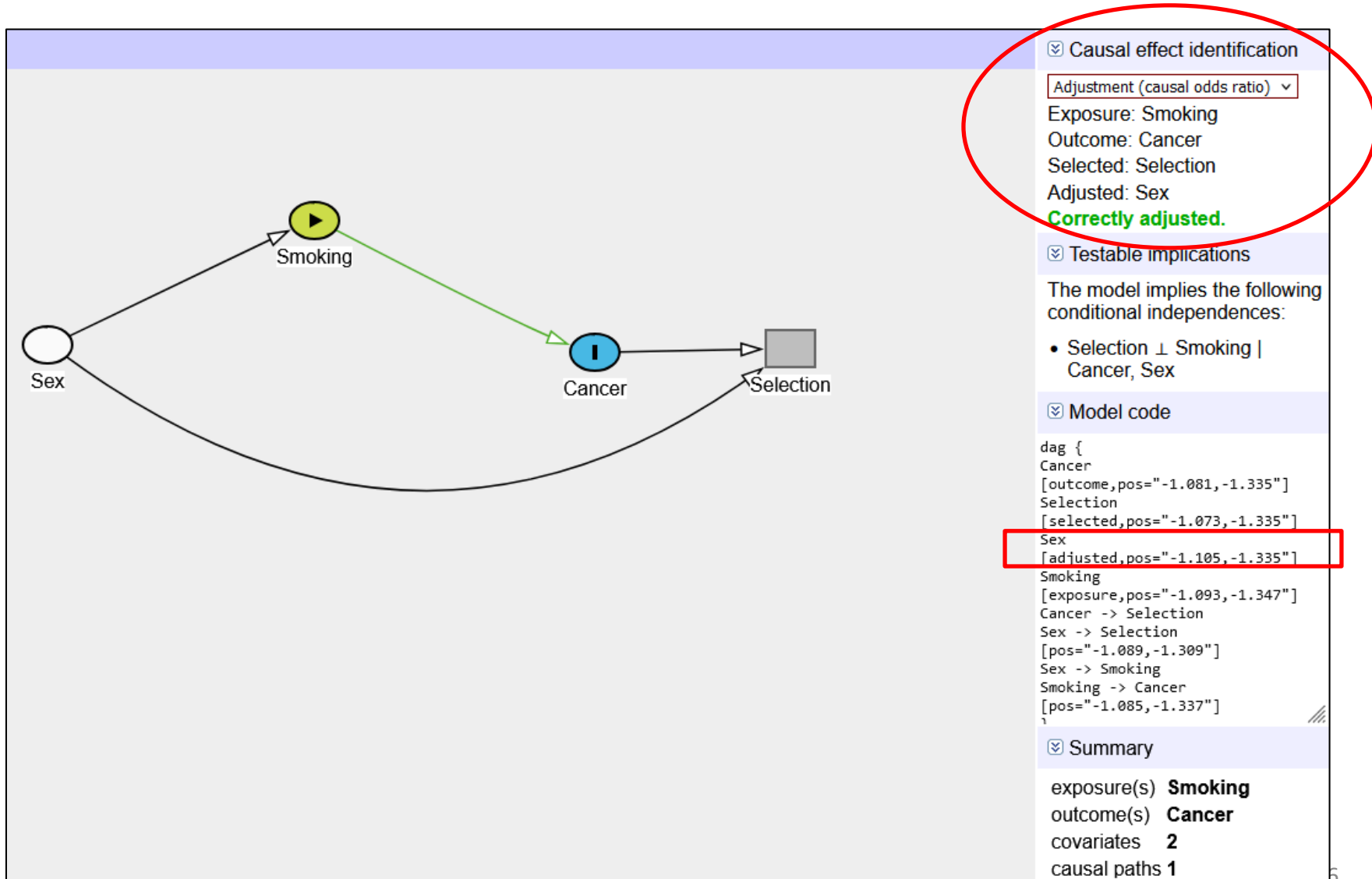


```
dag {  
  Cancer [outcome,pos="-1.078,-1.344"]  
  Selection [selected,pos="-1.073,-1.344"]  
  Sex [pos="-1.085,-1.344"]  
  Smoking [exposure,pos="-1.083,-1.346"]  
  Cancer -> Selection  
  Sex -> Selection [pos="-1.081,-1.341"]  
  Sex -> Smoking  
  Smoking -> Cancer [pos="-1.080,-1.345"]}
```

Look at the results

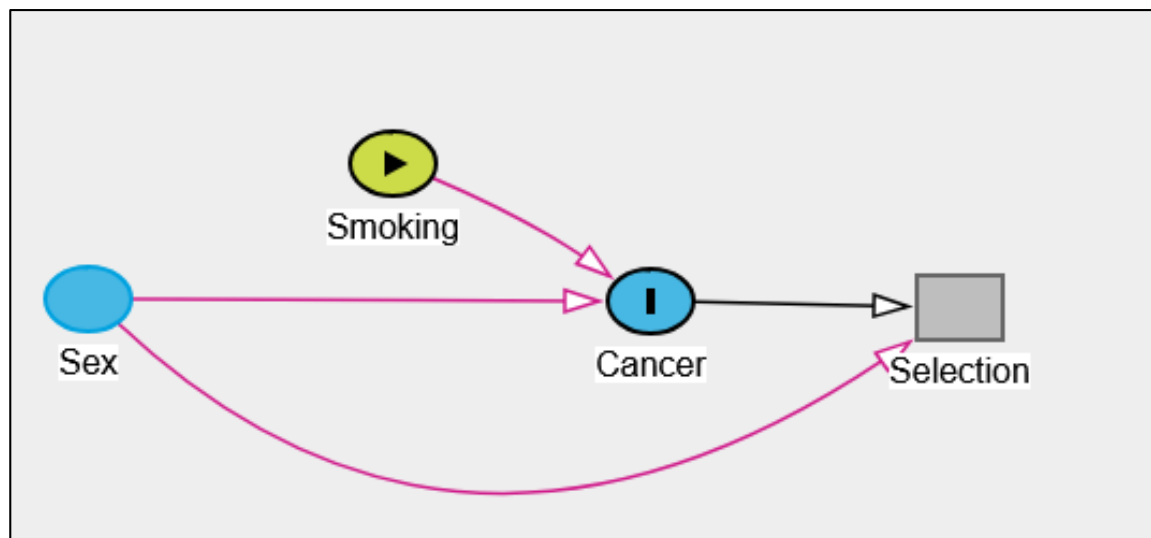


Need to adjust for the matching factor (sex)



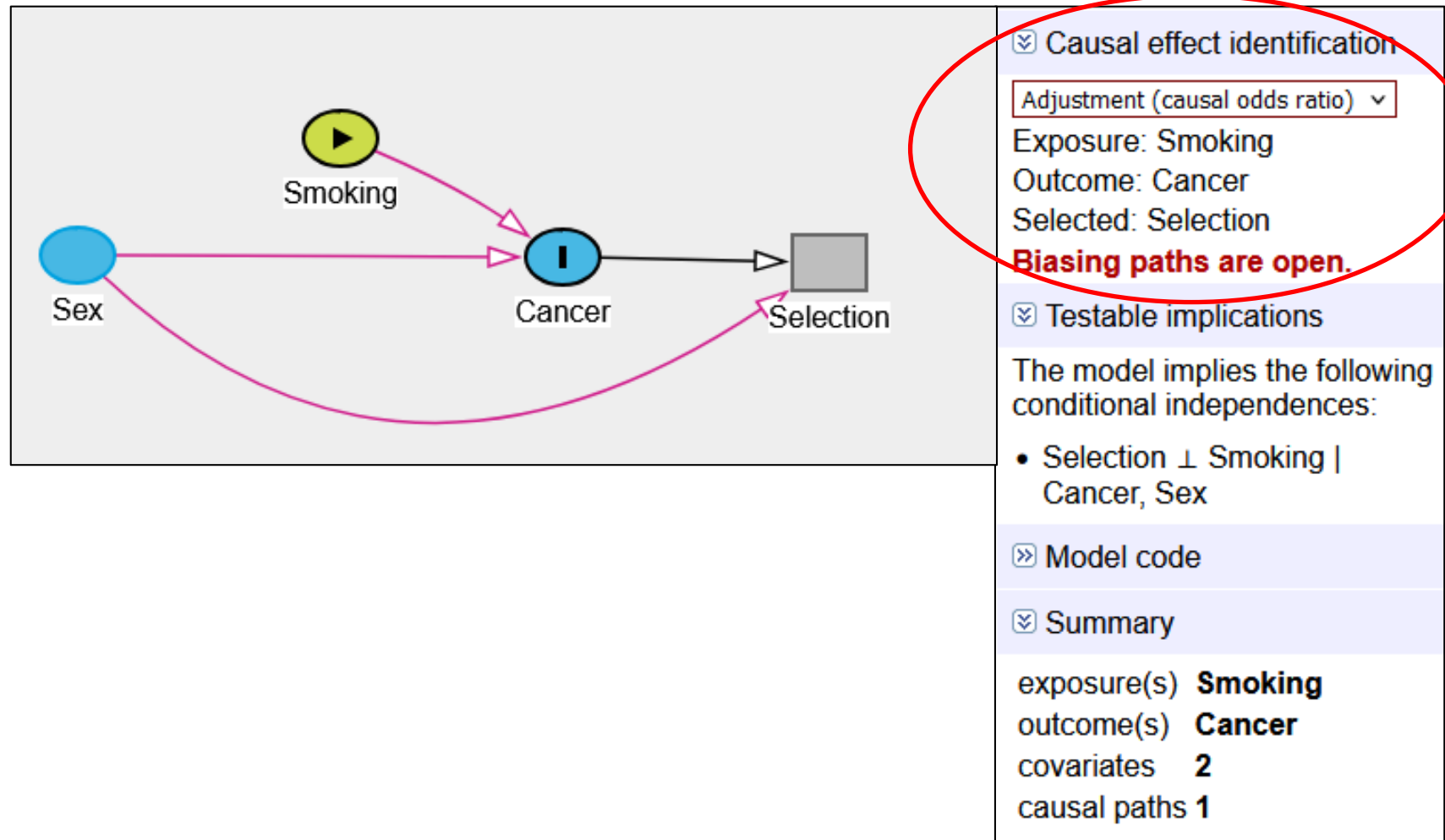
Example 6

A case-control study on smoking and lung cancer, matched on sex, with sex a cause of the outcome but not of exposure



```
dag {  
  Cancer [outcome,pos="-1.079,-1.343"]  
  Selection [selected,pos="-1.076,-1.343"]  
  Sex [pos="-1.085,-1.344"]  
  Smoking [exposure,pos="-1.082,-1.345"]  
  Cancer -> Selection  
  Sex -> Cancer  
  Sex -> Selection [pos="-1.081,-1.341"]  
  Smoking -> Cancer [pos="-1.080,-1.344"]}
```

Look at the results



Need to adjust for the matching factor (sex)

The image displays a causal diagram and a software interface for causal effect identification. The diagram shows three nodes: 'Sex' (white circle), 'Smoking' (yellow circle with a play button), and 'Cancer' (blue circle with an 'I'). A green arrow points from 'Smoking' to 'Cancer'. A black arrow points from 'Sex' to 'Cancer'. A black arrow points from 'Cancer' to 'Selection' (grey rectangle). A curved black arrow points from 'Sex' to 'Selection'.

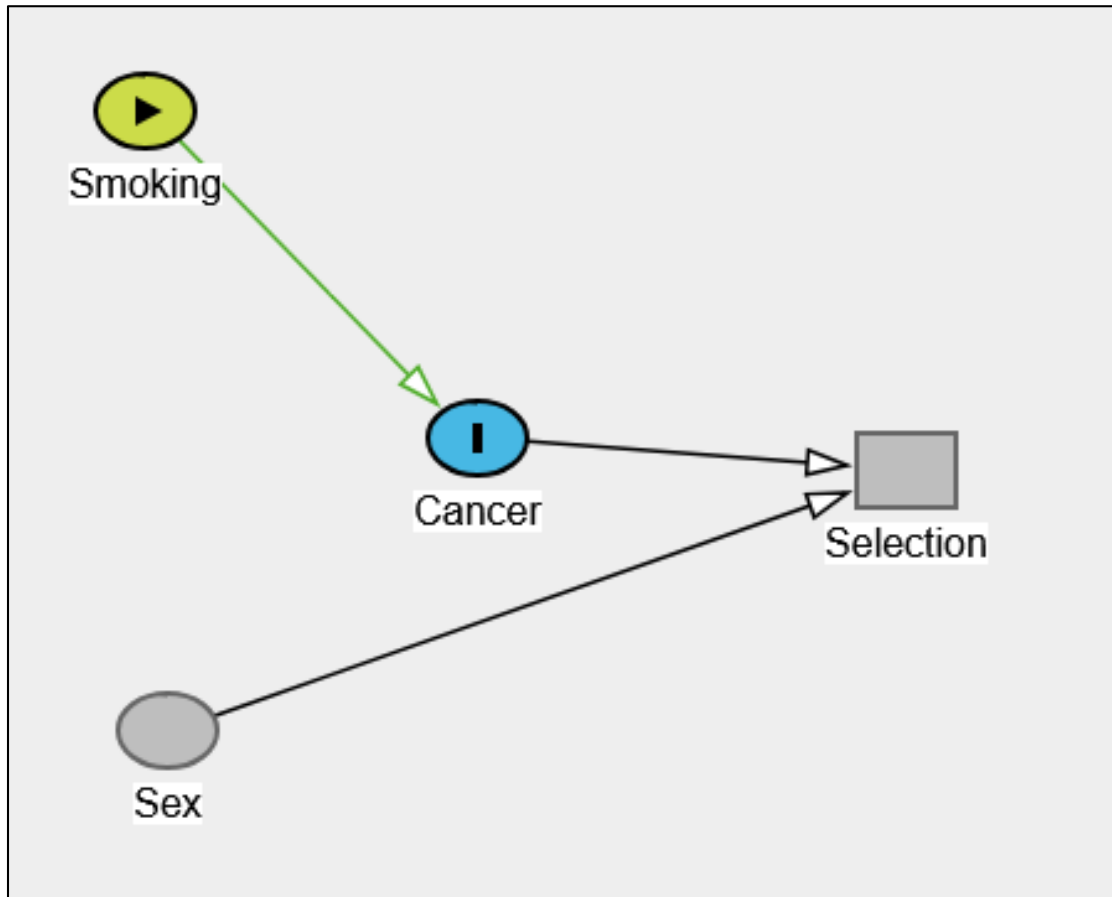
The software interface on the right has a red circle around the 'Causal effect identification' section. This section includes a dropdown menu set to 'Adjustment (causal odds ratio)', and the following text: 'Exposure: Smoking', 'Outcome: Cancer', 'Selected: Selection', 'Adjusted: Sex', and 'Correctly adjusted.' in green. Below this is the 'Testable implications' section, which states 'The model implies the following conditional independences:' and lists '• Selection \perp Smoking | Cancer, Sex'. The 'Model code' section shows a DAG with nodes Cancer, Selection, Sex, and Smoking, and their respective positions. The 'Sex' node's position is highlighted with a red box: '[adjusted,pos="-1.085,-1.344"]'. The 'Summary' section lists: 'exposure(s) Smoking', 'outcome(s) Cancer', 'covariates 2', and 'causal paths 1'.

```
dag {
  Cancer
  [outcome,pos="-1.079,-1.343"]
  Selection
  [selected,pos="-1.076,-1.343"]
  Sex
  [adjusted,pos="-1.085,-1.344"]
  Smoking
  [exposure,pos="-1.082,-1.347"]
  Cancer -> Selection
  Sex -> Cancer
  Sex -> Selection
  [pos="-1.081,-1.341"]
  Smoking -> Cancer
  [pos="-1.080,-1.344"]
}
```

exposure(s) **Smoking**
outcome(s) **Cancer**
covariates **2**
causal paths **1**

Example 7

A case-control study on smoking and lung cancer, matched on sex, with sex neither a cause of outcome nor exposure

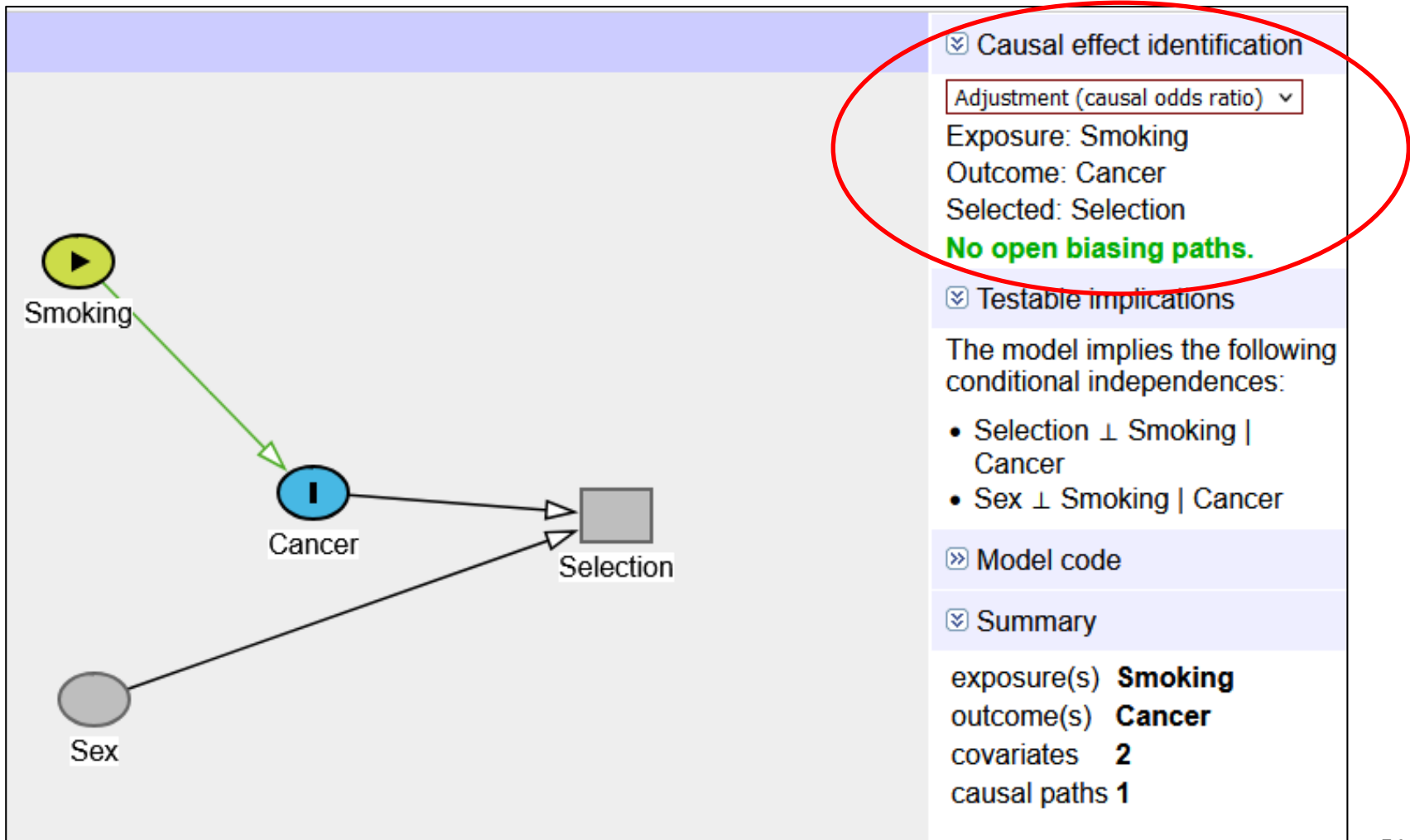


```
dag {  
  Cancer [outcome,pos="-1.080,-1.344"]  
  Selection [selected,pos="-1.076,-1.344"]  
  Sex [pos="-1.083,-1.342"]  
  Smoking [exposure,pos="-1.083,-1.346"]  
  Cancer -> Selection  
  Sex -> Selection  
  Smoking -> Cancer  
}
```

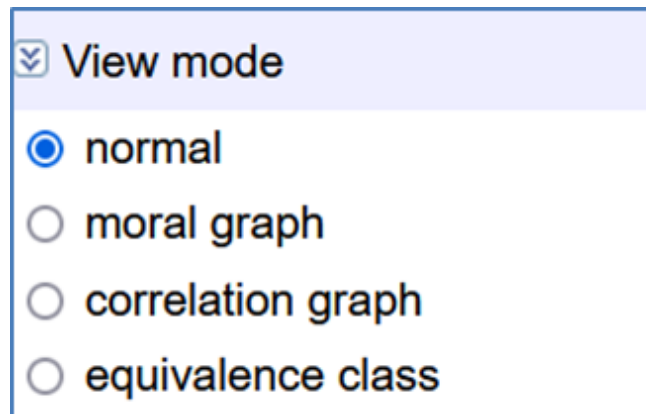
Look at the results

No need to adjust for matching factors

(but to no reason to have matched on them!)



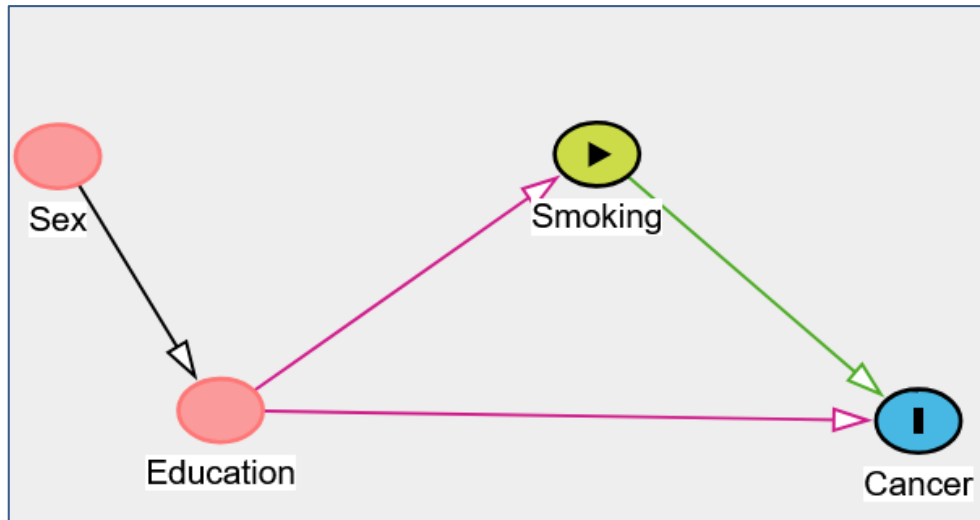
View mode of the graph



View mode

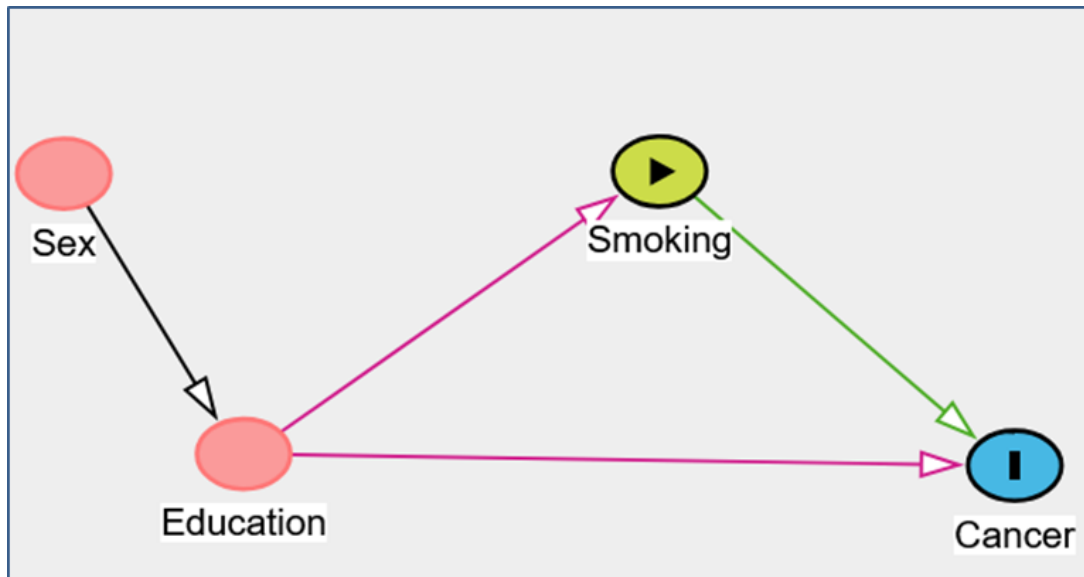
- ☒ normal
- ☐ moral graph
- ☐ correlation graph
- ☐ equivalence class

Example 8: Smoking and lung cancer



```
dag {  
  Cancer [outcome,pos="-0.895,-1.060"]  
  Education [pos="-1.629,-1.078"]  
  Sex [pos="-1.800,-1.527"]  
  Smoking [exposure,pos="-1.233,-1.531"]  
  Education -> Cancer  
  Education -> Smoking  
  Sex -> Education  
  Smoking -> Cancer  
}
```

Look at the results



☑ Causal effect identification

Adjustment (total effect) ▾

Exposure: Smoking

Outcome: Cancer

Biasing paths are open.

Minimal sufficient adjustment sets for estimating the total effect of Smoking on Cancer:

- Education

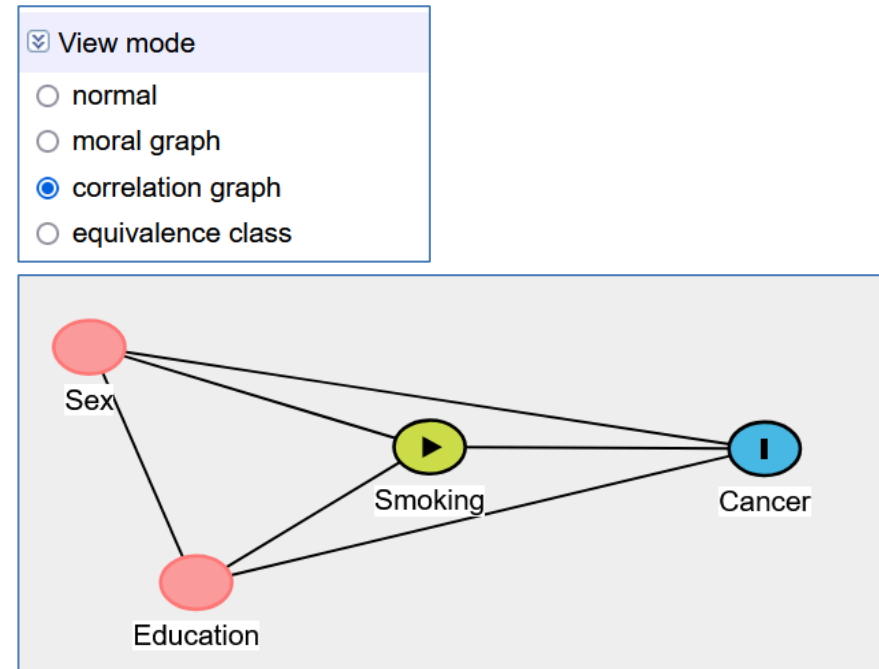
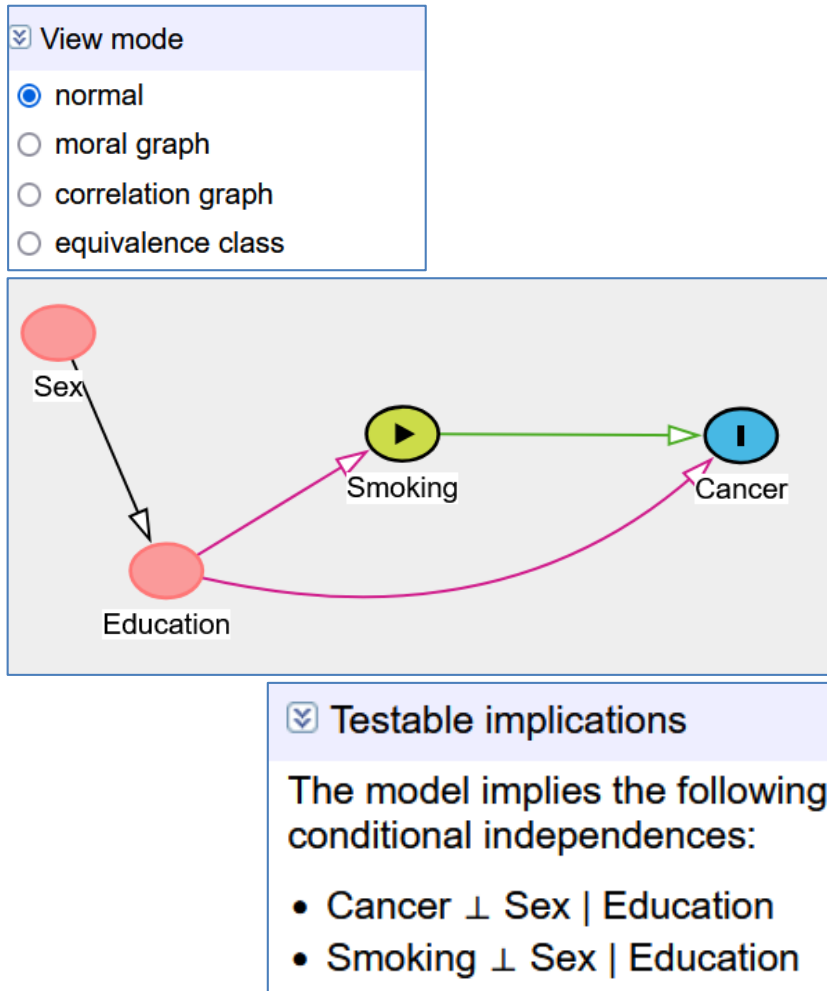
☑ Testable implications

The model implies the following conditional independences:

- $\text{Cancer} \perp \text{Sex} \mid \text{Education}$
- $\text{Smoking} \perp \text{Sex} \mid \text{Education}$

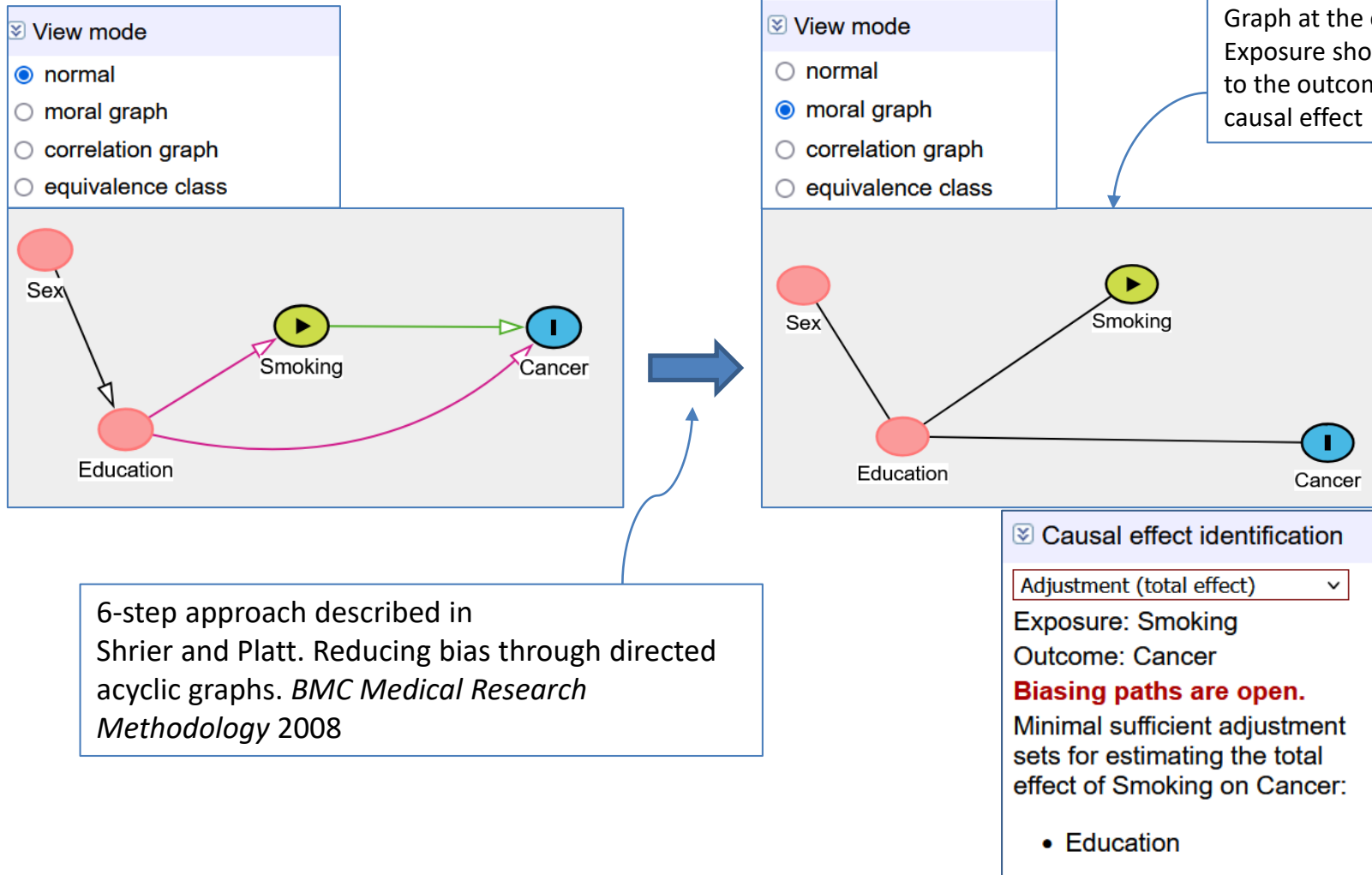
Correlation graph

- Not a DAG, but a simple graph with lines instead of arrows.
- Connects each pair of variables that, according to the DAG, could be statistically dependent.
- These pairwise independencies are also listed in the “Testable implications” field



Moral graph

- Undirected graph version of the DAG used by DAGitty to identify minimal sufficient adjustment sets
- Recommended to verify the calculation by hand for complicated DAG



DAGitty selected references

- Textor et al. [A Graphical Tool for Analyzing Causal Diagrams](#). *Epidemiology* 2011
- Textor et al. [Robust causal inference using DAGs: the R package 'dagitty'](#). *Int J Epidemiol* 2016
- Shrier et al. [Identifiability of causal effects in test-negative design studies](#). *Int J Epidemiol* 2023
- Shrier and Platt. [Reducing bias through directed acyclic graphs](#). *BMC Medical Research Methodology* 2008
- Textor. [Drawing and Analyzing Causal DAGs with DAGitty](#). DAGitty user manual Version July 18, 2023. <https://www.dagitty.net/manual-3.x.pdf>. Consulted on August 18, 2025