A Comparison of Different Techniques for Causal Inference Graphs and an Application to Obsessive-Compulsive Disorder

Jolanda Kossakowski

University of Amsterdam
J.J.Kossakowski@uva.nl
http://jolandakossakowski.eu

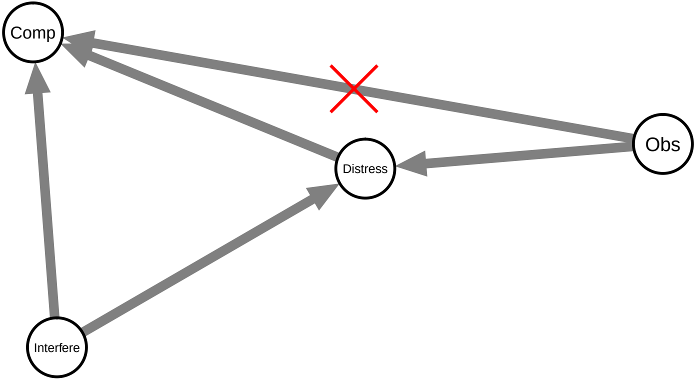
June 13, 2019

In Search for Causality

In psychology, we often want to know what the cause of a variable is

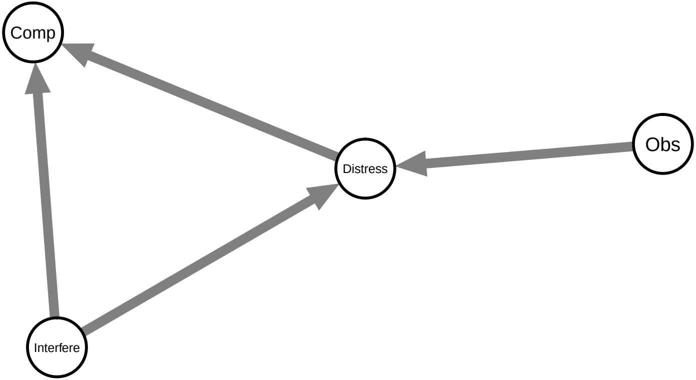
“a causal relation is a relation between two variables where, when changing one variable, we expect to observe a change in the other variable”

In the case of multiple variables, we want to determine whether a causal relation is direct or indirect



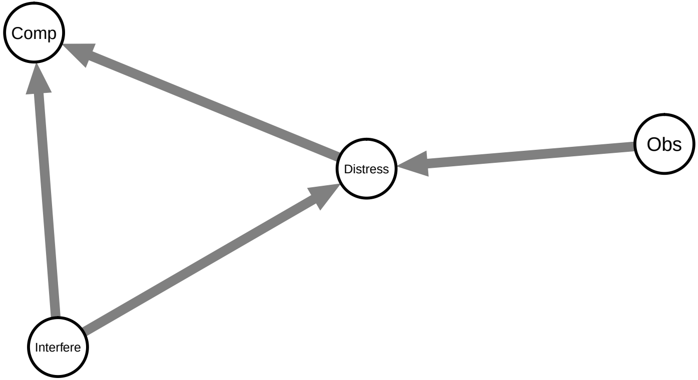
In Search for Causality

In the case of multiple variables, we want to determine whether a causal relation is direct or indirect



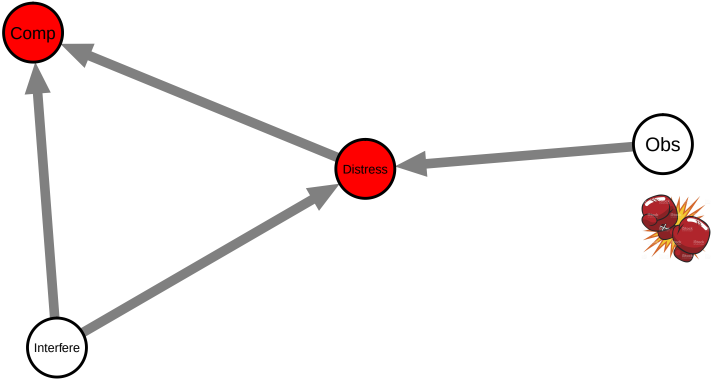
From Observational to Experimental Data

Observational data



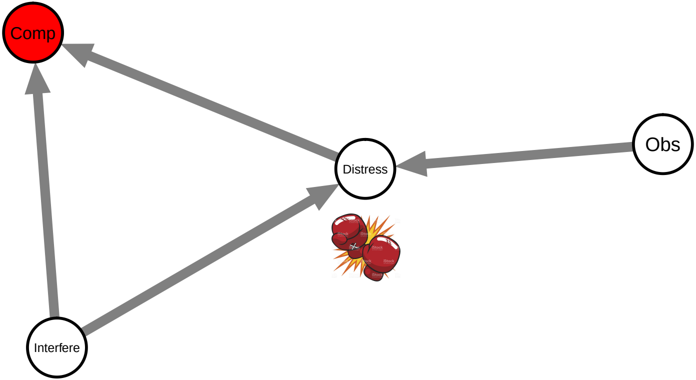
From Observational to Experimental Data

Experimental data



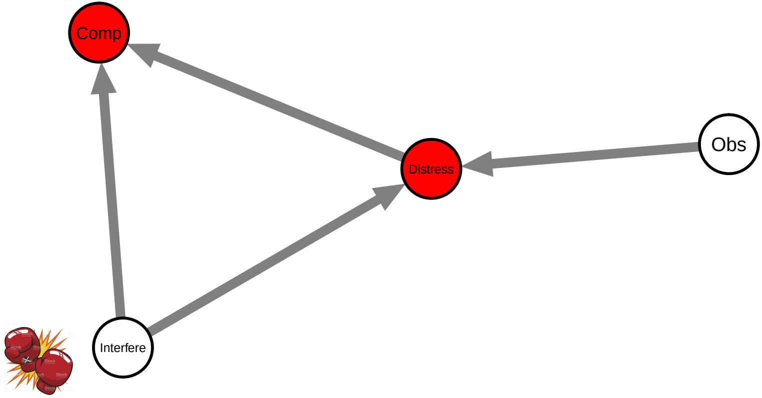
From Observational to Experimental Data

Experimental data



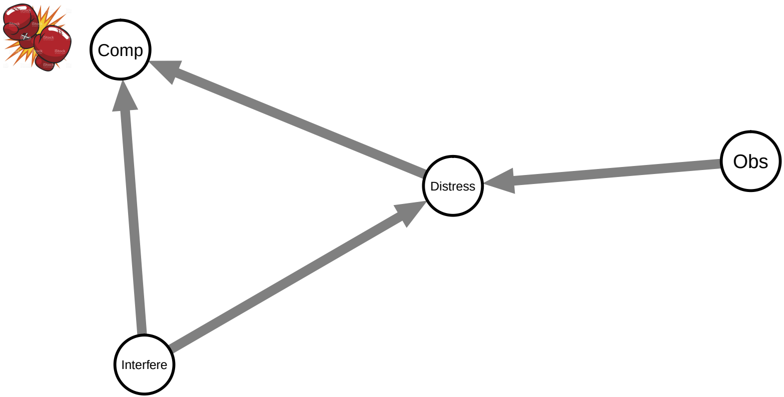
From Observational to Experimental Data

Experimental data



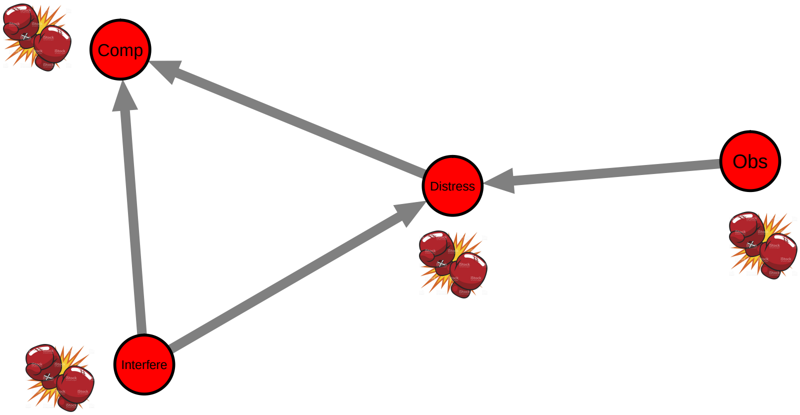
From Observational to Experimental Data

Experimental data



From Observational to Experimental Data

Experimental data



Goal

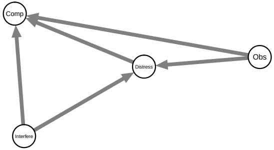
- The goal of this study was to investigate different options to combine data from different sources (observational and experimental).
- We looked at five different techniques:
    - Peter and Clark (PC-) algorithm
    - Downward Ranking of Feed-Forward Loops (DR-FFL)
    - Transitive Reduction for Weighted Signed Digraphs (TRANSWESD)
    - Invariant Causal Prediction (ICP-) algorithm
    - Hidden Invariant Causal Prediction (HICP-) algorithm

Methods for Causal Inference

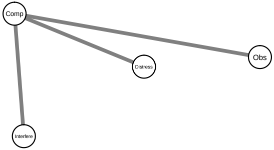
PC-algorithm[1]

- Step 1: Determine skeleton graph by means of conditional (in)dependence
- Step 2: Determine causal graph by finding collider structures

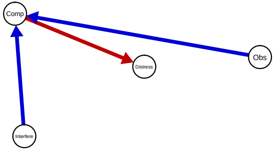
True graph



Step 1



Step 2

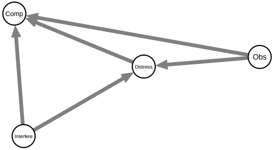


Methods for Causal Inference

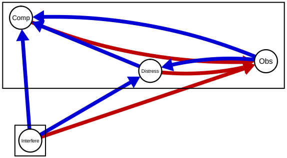
DR-FFL algorithm[2]

- Step 1: Generate a perturbation graph that shows all causal relations $> \beta$
- Step 2: Determine strongly connected components (SCCs)
- Step 3: Remove edges between SCCs if an alternative path exists between them

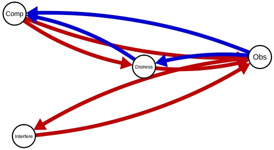
True graph



Step 1 & 2



Step 3

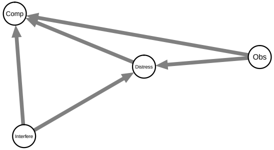


Methods for Causal Inference

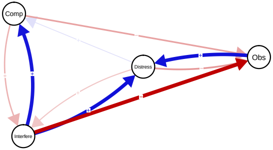
TRANSWESD-algorithm[3]

- Step 1: Generate a perturbation graph that shows all causal relations $> \beta$ and an absolute change $> \gamma$
- Step 2: Remove edges between if an alternative path exists without a cycle, a sign equal to the original and a maximum weight $< e_{ij} \cdot \alpha$

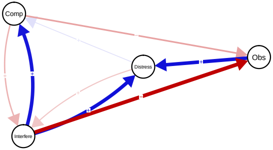
True graph



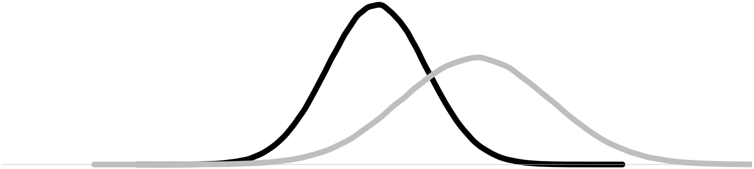
Step 1



Step 2

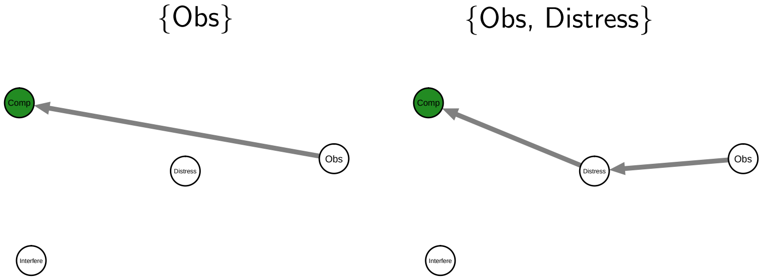


- Invariant Causal Prediction (ICP) regresses one *target* variable on each possible subset of remaining variables.
- Per regression, the regression coefficients and residual distributions are tested for equality ("invariance") across time points.

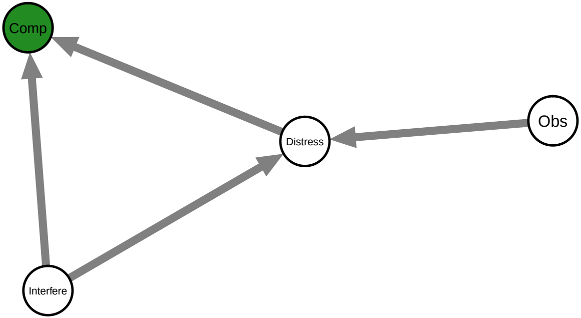


- We only draw directed edges if one or more nodes show invariance across **all** time points.
- This process is repeated for all nodes in the graph.
- Directed edges represent unweighted causal relations

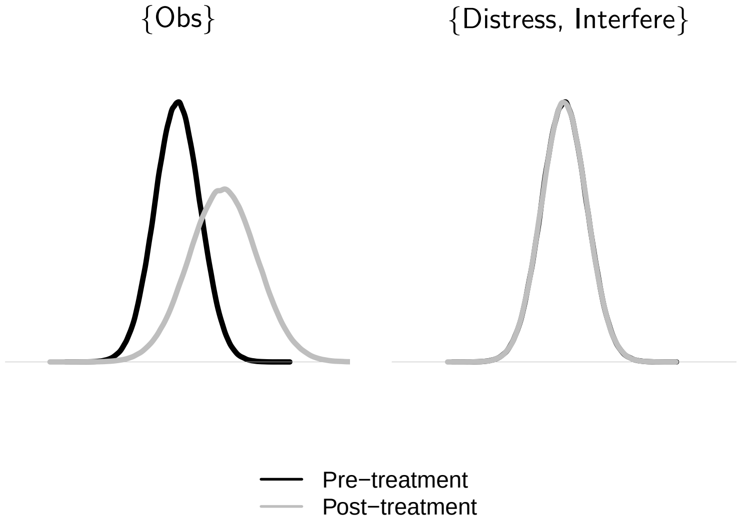
The need for all possible subsets



If we only look at single variables, we may find direct effects that are truly indirect ones.



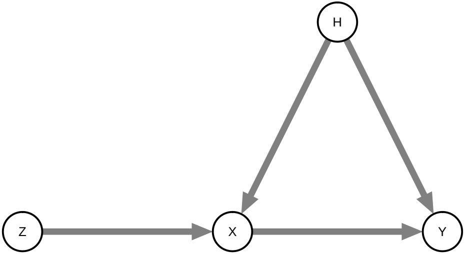
No causal relations	$\{\emptyset\}$
1 causal relation	{Obs} {Distress} {Interfere}
2 causal relations	{Obs, Distress} {Obs, Interfere} {Distress, Interfere}
3 causal relations	{Obs, Distress, Interfere}


{Obs}
{Distress, Interfere}
Pre−treatment
Post−treatment

Methods for Causal Inference

HICP-algorithm[5]

The ICP-algorithm assumes that the causal coefficients and its residuals are uncorrelated.



This assumption cannot be satisfied when hidden variables are present

- The hidden ICP-algorithm searches for the set of causes for a target variable, while controlling for “hidden variables”
    - Hidden variables are variables that are not measured, but may affect variables that were measured
- The algorithm partials out the effect of the hidden variable by means of an “instrumental variable”

Simulation study

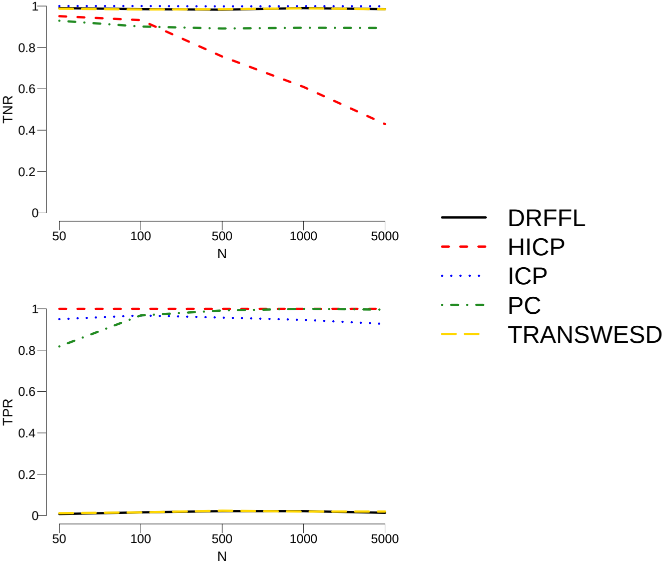
We evaluated the performance of each algorithm in a simulation where we varied the following parameters:

- Number of nodes $V = \{5, 10\}$
- Graph density $d = \{0.1, 0.25, 0.5\}$
- Sample size $n = \{50, 100, 500, 1000, 5000\}$
- Perturbation mean $\overline{m} = \{1, 5\}$
- Perturbation standard deviation $sd = \{0.5, 5\}$
- $\beta = \{0.5, 1, 1.64, 1.96, 2.58\}$

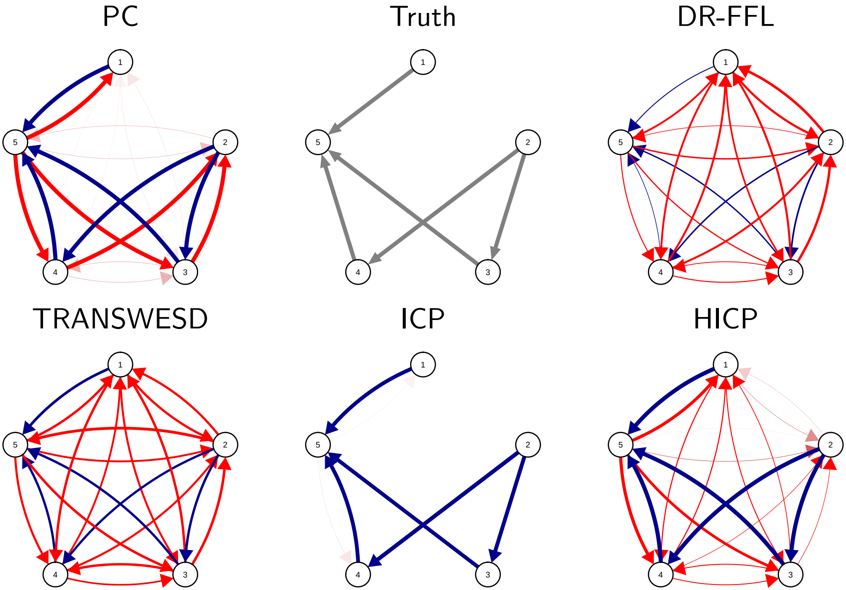
Simulation results

Specificity

Sensitivity



Simulation results



Estimating a Causal Graph for OCD

- We applied both the ICP and the HICP algorithm on an empirical dataset
- We combined these results with a literature study
  - We searched the literature for evidence for specific causal relations between nodes
  - Search happened in a non-structural manner
  - Text was searched for OCD symptoms and verbs that may indicate a causal relation
  - i.e., "Obsessive-compulsive disorder (OCD) is a common anxiety disorder characterized by intrusive thoughts that are difficult to dismiss and that increase anxiety" (Marker, Calamari, Woodard, & Riemann, (2006))
  - All evidence was double checked by an independent researcher.

Estimating a Causal Graph for OCD

Measure

Patients completed the *Yale-Brown Obsessive-Compulsive Scale* (Y-BOCS) at each measurement.

Obsessions	Item	Compulsions	Scale
O_Time	Time occupied by	C_Time	$0-4$
O_Int	Interference due to	C_Int	$0-4$
O_Dist	Distress associated with	C_Dist	$0-4$
O_Res	Resistance against	C_Res	$0-4$
O_Cont	Degree of control over	C_Cont	$0-4$

Estimating a Causal Graph for OCD

Empirical data

Invited
$N = 3474$

Baseline $N = 3121$ → **Progress1** $N = 1909$ → **Progress2** $N = 1308$ → **Post** $N = 2617$

16.65 days | 15.80 days | 23.25 days

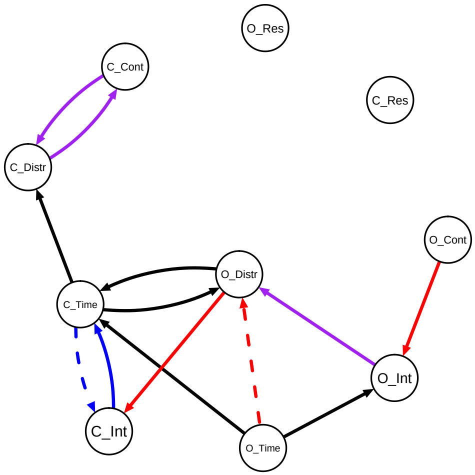
Included
$N = 903$

Estimating a Causal Graph for OCD

Patient Characteristics

- 52.93% female
- Mean age: 29.69 ($SD = 11.42$)
- Average treatment time: 56.39 days ($SD = 20.29$)
- Average Y-BOCS score significantly decreased over time $F(3, 2706) = 673.85, p < .001, \eta^2 = 0.16$

Estimating a Causal Graph for OCD



Red: ICP

Blue: HiddenICP

Purple: Combination

Black: Literature

Solid: One method

Dashed: Both algorithm and literature

Conclusion

- By combining different sources, we can infer causal relations, and possible cyclic patterns
- This causal graph may give us different and novel insights into the structure of OCD
- More research has to be conducted to improve both the algorithms and design of the literature study

Contact

Kossakowski, J.J., Waldorp, L. J., & van der Maas, H. L. J. (2019). The Race for Causality: A Comparison of Different Techniques for Causal Inference Graphs. *In preparation*.

Kossakowski, J. J., McNally, R. J., van Oudheusden, L., Riemann, B. C., Waldorp, L. J., & van der Maas, H. L. J. (2019). A Causal Graph of Obsessive-Compulsive Disorder. *In preparation.*

Contact info:
`http://jolandakossakowski.eu`
`J.J.Kossakowski@uva.nl`

References

[1]Spirtes, Glymour, & Scheines, (2000). Causation, Prediction, and Search.
[2]Pinna et al., (2013). Reconstruction of large-scale regulatory networks based on perturbation graphs and transitive reduction: improved methods and their evaluation.
[3]Klamt, et al., (2010). TRANSWESD: inferring cellular networks with transitive reduction.
[4]Meinshausen et al. (2016) Methods for causal inference from gene perturbation experiments and validation.
[5]Peters, Janzing, & Schölkopf, (2017). Elements of Causal Inference: Foundations and Learning Algorithms.