Young statistical research in Denmark

November 17, 2021

YOUNG STATISTICIANS DENMARK

Young Statisticians Denmark (YSD)

A society that plans social and scientific events, especially for students and young professionals working with statistics.

- Share knowledge in a relaxed atmosphere.
- "Young" and "Statistician" is broadly defined.
- Part of the Danish Society for Theoretical Statistics.

YOUNG
STATISTICIANS
DENMARK

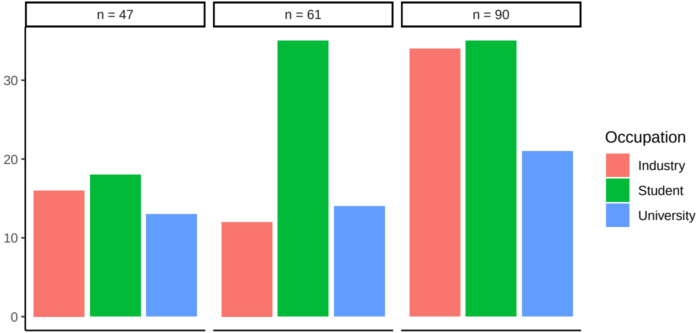
Previous events

22 events since 2015.

Neurobiology meets Statistics
Science talk and pub quiz
Career event
Talking about the p-word
...

YOUNG
STATISTICIANS
DENMARK

Crowd


n = 47
n = 61
n = 90
30
20
10
0
Occupation
Industry
Student
University

Upcoming events

2 upcoming events within the next two months.

November 23 Statisticians in the wild (vol. 2)
January 13 Event on causality

Young statistical research in Denmark

What questions occupy the minds of young statistical researchers in Denmark? What topics do they spend their time on studying?

YSD presents:

Sneha Das

- Postdoc at the Section of Statistics and Data Analysis (The Technical University of Denmark)
- PhD from Aalto University Finland (defense in one week)

YOUNG
STATISTICIANS
DENMARK

Multimodal Signal Modelling for Intervening and Managing Mental Disorders

Sneha Das

Statistics and Data Analysis, Technical University of Denmark (DTU)

Outline

- Motivation
- General introduction
- Audio
- Biosignals
- Conclusions

Mental Health and Mental Disorders

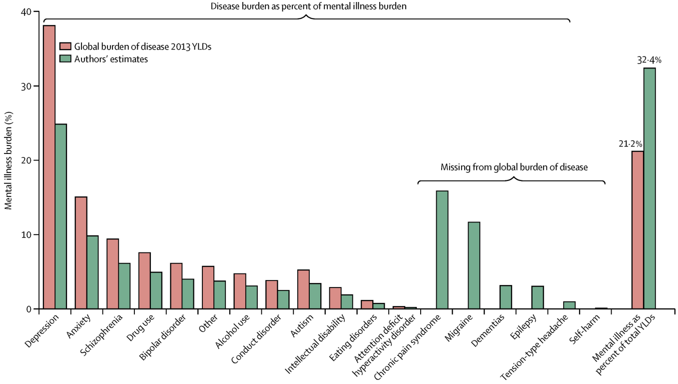


Figure: Original image from Vigo et al., 2016

- Mental illness is one of the leading causes of global disease burden (Prince et al., 2007; Vigo et al., 2016).
- In Denmark, 15% of youth will be diagnosed with a psychiatric disorder before their 18th birthday (Dalsgaard et al., 2020).

WristAngel: Intervention and Research for OCD Treatment I

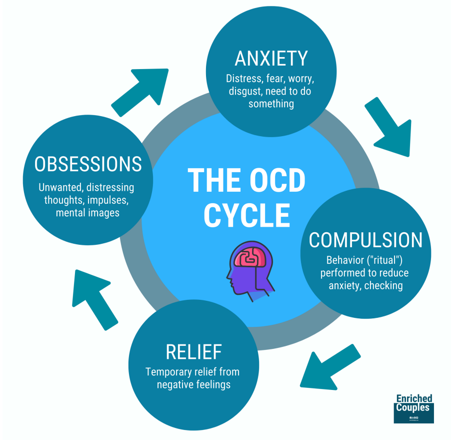


Figure: Obsessions and compulsions behave cyclically. Original image from `https://medium.com/amalgam/ocd-is-not-what-you-think-it-is-ee818028e79c`

- Mental disorder wherein "People are caught in a cycle of obsession and compulsions".
- Obsessions → intrusive and disruptive urges, thoughts, images, etc.
- Compulsions → behavior to overcome obsessions, distress.
- In 2010, anxiety disorders - including obsessive-compulsive disorders -alone cost Europe over €74 billion (Gustavsson et al., 2011).

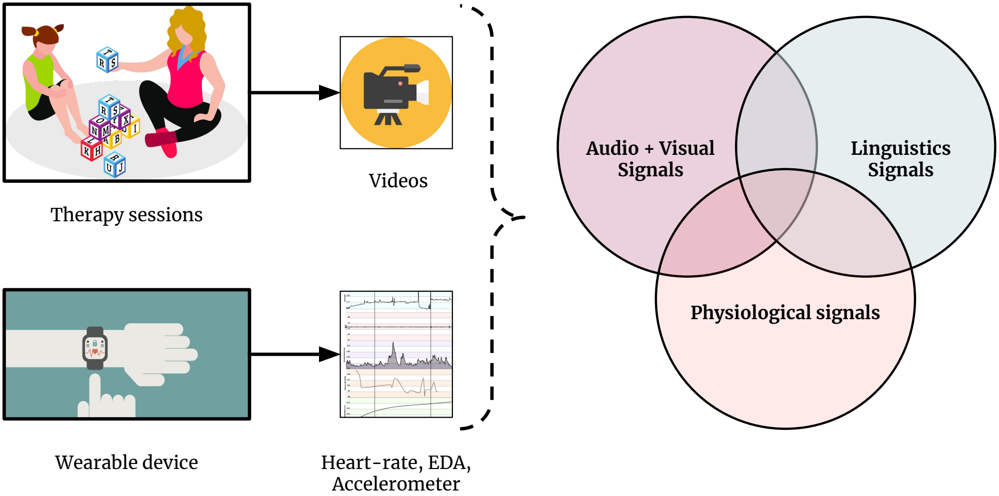
WristAngel: Intervention and Research for OCD Treatment II

Identify and predict impending OCD events and provide useful interventions $\rightarrow$ progression and severity of disorder.
Aid in delivering cognitive behavioral therapy to patients.

WristAngel

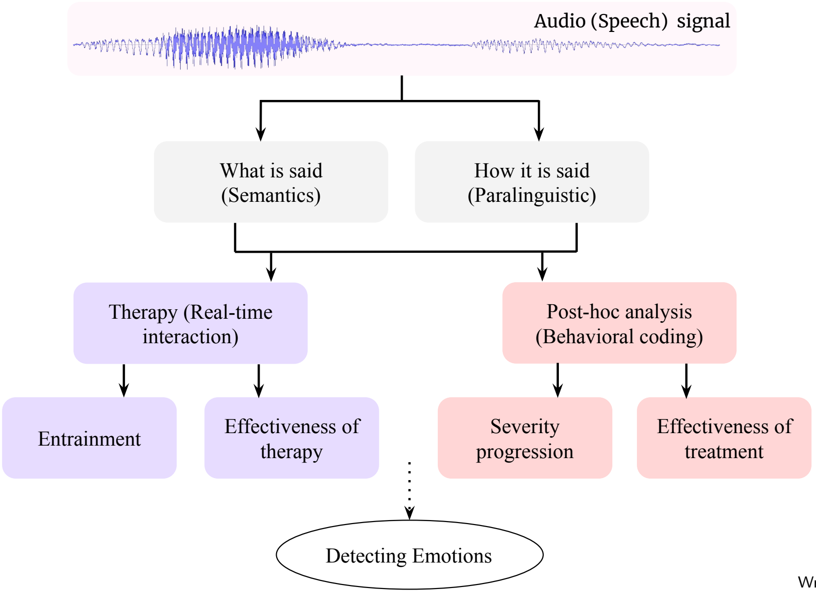
• Sneha Das	DTU Compute
• Line H. Clemmensen	DTU Compute
• Nicole Nadine Lønfeldt	Child and Adolescent Mental Health Center, KU Hospital
• Anne Katrine Pagsberg	Faculty of Health, Department of Clinical Medicine, KU
• Nicklas Leander Lund	DTU Compute

Data and Signals



AUDIO SIGNALS

Role of Audio (Speech) in OCD Treatment



Speech Emotion Detection I

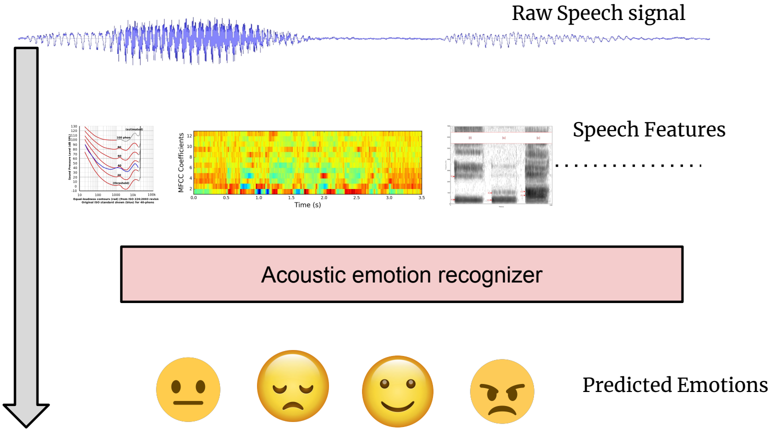


Figure: Image sources https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd; https://commons.wikimedia.org/wiki/File:Lindos1.svg; https://commons.wikimedia.org/wiki/File:Spectrogram_-iua-.png

Speech Emotion Detection II

Conventional approaches

• Statistical ML and signal processing	HMM, GMM, SVM
• Deep learning (DL)	RNN, CNN, LSTM with deep architectures
• Hybrid	Eg., DL +SVM

Persistent challenges

• Generalization	corpora, languages $\rightarrow$ cultural, phonetic differences
• Low-resource corpora	Small data set and lack of labels
• Black-boxes	

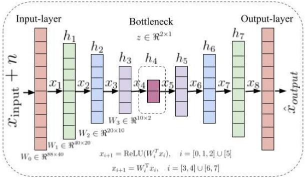
Methodology

Transferable feature embedding

Denoising autoencoder (DAE)

$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{Rec}}$

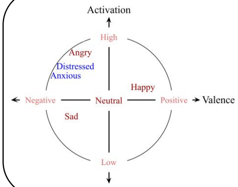
Discriminating subspace



Semi-supervision (metric learning)

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{Rec}} + \mathcal{L}_{\text{Met}}$$

Emotion relevant embedding



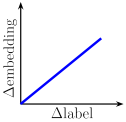
Dimensional emotion model (Activation, valence)

1. Consistent over languages, in contrast to emotion class labels
2. Allows to model emotion classes that does not exist in training dataset.

Continuous metric loss for emotion modelling

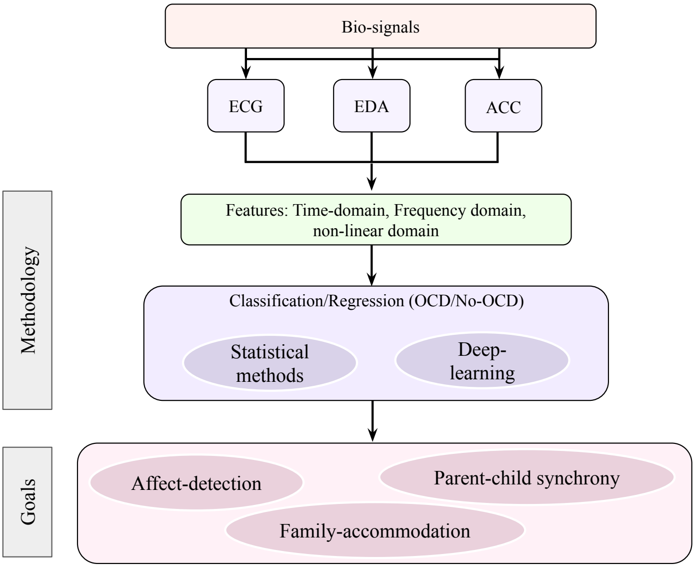
$$\mathcal{L}_{\text{Met}} = \mathcal{L}_{\text{Slope}} + \mathcal{L}_{\text{Res}}$$

Label: Activation, valence

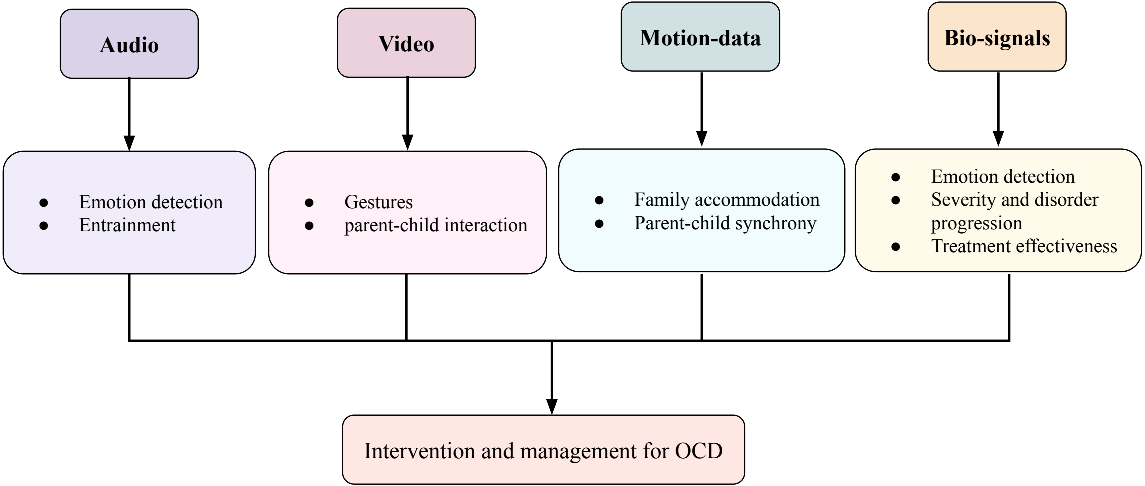


BIOSIGNALS

Signals, Methods and Goals



Tying Modalities Together



YSD presents:

Nikolaj Thams

- PhD student at Copenhagen Causality Lab, MATH (University of Copenhagen)
- Master in Statistics from University of Copenhagen

YOUNG STATISTICIANS DENMARK

Causality for Distributional Robustness

DSTS 50 year anniversary

Nikolaj Thams
PhD Student at MATH, Univ. Copenhagen

Using Causality for Distributional Robustness

- The road from full to partial exogeneity
- Some detours about our work

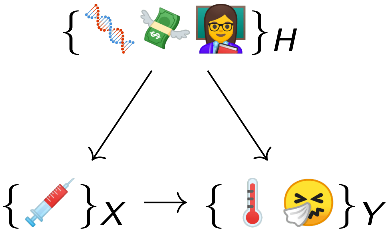
Perfect exogeneity: Randomized experiments

Goal: Learn effect X on Y.

$$\{💉\}_X \longrightarrow \{🌡️ 🤧\}_Y$$

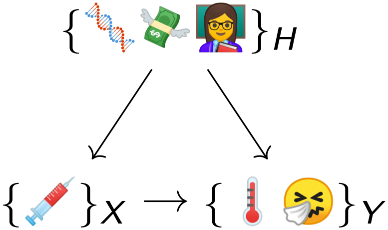
Perfect exogeneity: Randomized experiments

Goal: Learn effect X on Y.
Problem: Latent confounding factors H.



Perfect exogeneity: Randomized experiments

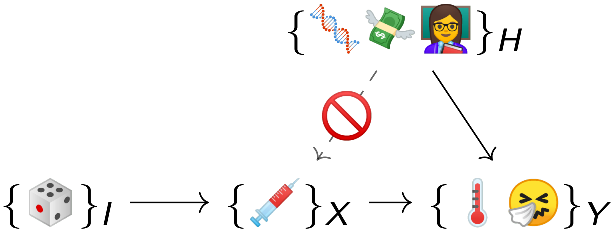
Goal: Learn effect X on Y.
Problem: Latent confounding factors H.
Solution: Use randomization to break confounding.



Why care about causal instead of confounded effect? Generalization!

Perfect exogeneity: Randomized experiments

Goal: Learn effect X on Y.
Problem: Latent confounding factors H.
Solution: Use randomization to break confounding.



Why care about causal instead of confounded effect? Generalization!

Enough exogeneity: Instrumental variables

We can't always randomize:

- Allocation unethical (e.g. smoking 🚬 or seatbelts 🔥🏎️)
- Interventions hard to define (e.g. data is pixels of image 🖼️)

Enough exogeneity: Instrumental variables

We can't always randomize:

- Allocation unethical (e.g. smoking 🚬 or seatbelts 🔥🏎️)
- Interventions hard to define (e.g. data is pixels of image 🖼️)

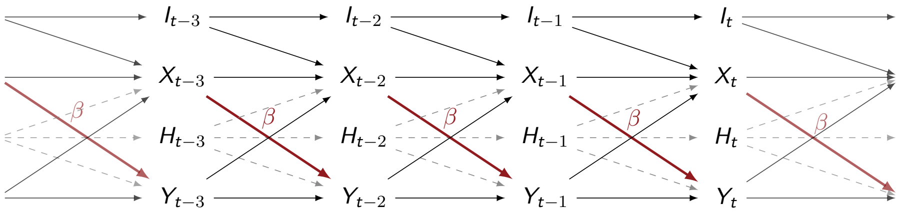
Instead: Instrumental variable (IV): I 1) independent of H and 2) only indirect effect on Y

$$\begin{tikzcd} & & H \arrow[dl] \arrow[dr] & \\ I \arrow[r] & X \arrow[rr] & & Y \end{tikzcd}$$

OLS regression: $Y - X\hat{\beta} \perp\!\!\!\perp X \longrightarrow$ IV regression: $Y - X\hat{\beta} \perp\!\!\!\perp I$.

$\rightarrow$ Causal effect β can still be estimated (Imbens and Rubin 2015) under assumptions, e.g. high rank I.

Detour 1: Instruments in time series



Currently, we are working on methods in systems with memory

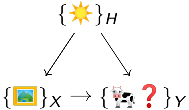
- Proper adjustment for memory
- Using also lagged instruments as instruments, to increase rank of instrument,

(Thams et al. 2021).

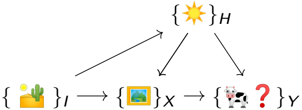
Some exogeneity: Invariance

$\{\text{🖼️}\}_X \to \{\text{🐄❓}\}_Y$

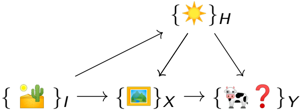
Some exogeneity: Invariance



Some exogeneity: Invariance

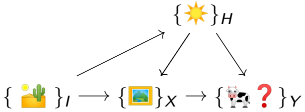


Some exogeneity: Invariance



- Causal solution: $Y - X\beta \perp\!\!\!\perp I$. IV require $\dim(I) \geq \dim(X)$. Image data: $\dim(X) = 1024^2$.

Some exogeneity: Invariance

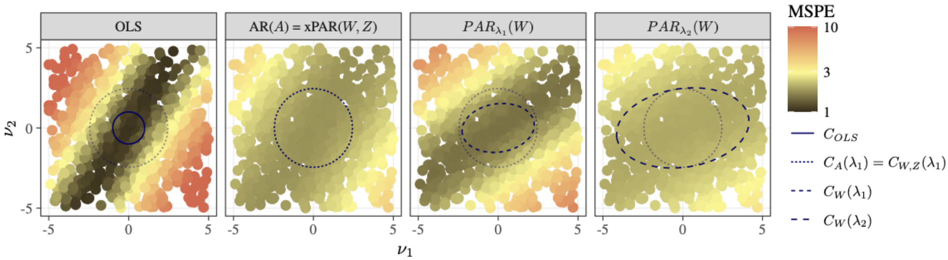


- Causal solution: $Y - X\beta \perp\!\!\!\perp I$. IV require $\dim(I) \geq \dim(X)$.
  Image data: $\dim(X) = 1024^2$.
- Instead, we can penalize our regression

$$\hat{\beta} \in \arg\min_{\beta} \|Y - X\beta\|^2 + \lambda \|\operatorname{cov}(Y - X\beta, I)\|^2,$$

Rothenhäusler et al. 2021 show that $\hat{\beta}$ generalizes well (but less than with full causal knowledge).

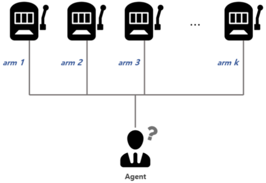
Detour 2: Generalization with proxy variables

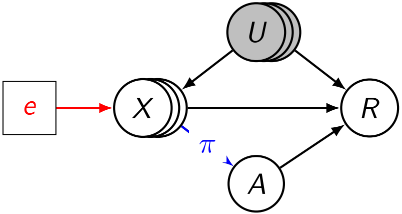


In Oberst et al. 2021, we show that

- If I is unobserved, but we have a noisy measurement (proxy) W of I, we still obtain generalization, although less.
- If we observe two proxies W, Z of I, we generalize as well as with I itself.

Detour 3: Using environments to generalize in decision making





Contextual (= covariates) + Bandits

In Saengkyongam et al. 2021,

- We develop invariant policies, that is policies that do not use confounded features.
- To do so, we use exogeneous environments, e.g. hospitals.
- Under assumptions, we generalize to new environments.

Conclusions

- If strong exogeneity is present, we can estimate causal effects
- If weak exogeneity is present, we can not estimate causal effects, but still get methods that generalize better
- We considered linear effects, but much can be generalized to non-linear functions (Christiansen et al. 2021)

References I

Christiansen, Rune, Niklas Pfister, Martin Emil Jakobsen, Nicola Gnecco, and Jonas Peters (2021). "A causal framework for distribution generalization". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1.

Imbens, Guido W and Donald B Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Oberst, Michael, Nikolaj Thams, Jonas Peters, and David Sontag (2021). "Regularizing towards Causal Invariance: Linear Models with Proxies". In: *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. Proceedings of Machine Learning Research. PMLR, pp. 8260–8270.

Rothenhäusler, Dominik, Nicolai Meinshausen, Peter Bühlmann, and Jonas Peters (2021). "Anchor regression: Heterogeneous data meet causality". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 83.2, pp. 215–246.

Saengkyongam, Sorawit, Nikolaj Thams, Jonas Peters, and Niklas Pfister (2021). "Invariant Policy Learning: A Causal Perspective". In: *arXiv preprint arXiv:2106.00808*.

References II

Thams, Nikolaj, Rikke Nielsen, Sebastian Weichwald, and Jonas Peters (2021). "Instrumental Time Series: Correcting for the Past, Identifiability, and Learning". In: *Work in progress.*

Questions?