Educational Course: *Global Open Science Electrophysiology*

OHBM 2024
JUNE 23~27, SEOUL, KOREA

Outlook and IT infrastructure

1. Overview of the GBC (Global Brain Consortium) strategy for Electrophysiology (Alan Evans)
2. The CBRAIN and LORIS neuroinformatics ecosystem (Bryan Caron)
3. Hierarchical Event Description (HED) annotation (Scott Makeig)
4. The EEGNet Data and Analytics Platform (Christine Rogers)
5. The Chinese WeBrain interface to GBC (Yun Qin)

qEEG and its applications

6. Introduction of qEEG (Pedro Valdes-Sosa)
7. Formulating clinical applications using qEEG (Maria Luisa Bringas-Vega)
8. QEEG toolbox (Jorge Bosch-Bayard)
9. **Causality and mediation tools for q-EEG and clinical applications** (Qing Wang (Vincent))

GLOBAL BRAIN CONSORTIUM

cbrain LORIS

EEGNet

HED

CONP PCNO

NeuroHub

WeBrain

jupyter

R Studio

python

OHBM 2024

International Federation of CLINICAL NEUROPHYSIOLOGY

incf

Educational Course: *Global Open Science Electrophysiology*

9. Causality and mediation tools for q-EEG and clinical applications

Qing Wang (Vincent)

https://github.com/Vincent-wq

vincent.w.qing@gmail.com

Shanghai Mental Health Center, School of Medicine,
Shanghai Jiao Tong University, Shanghai, China

Qing Wang (Vincent), Assistant Researcher, Shanghai Mental Health Center, School of Medicine, Shanghai Jiao Tong University

I have no disclosures.

Course materials: https://github.com/Vincent-wq/causal_course_eeg

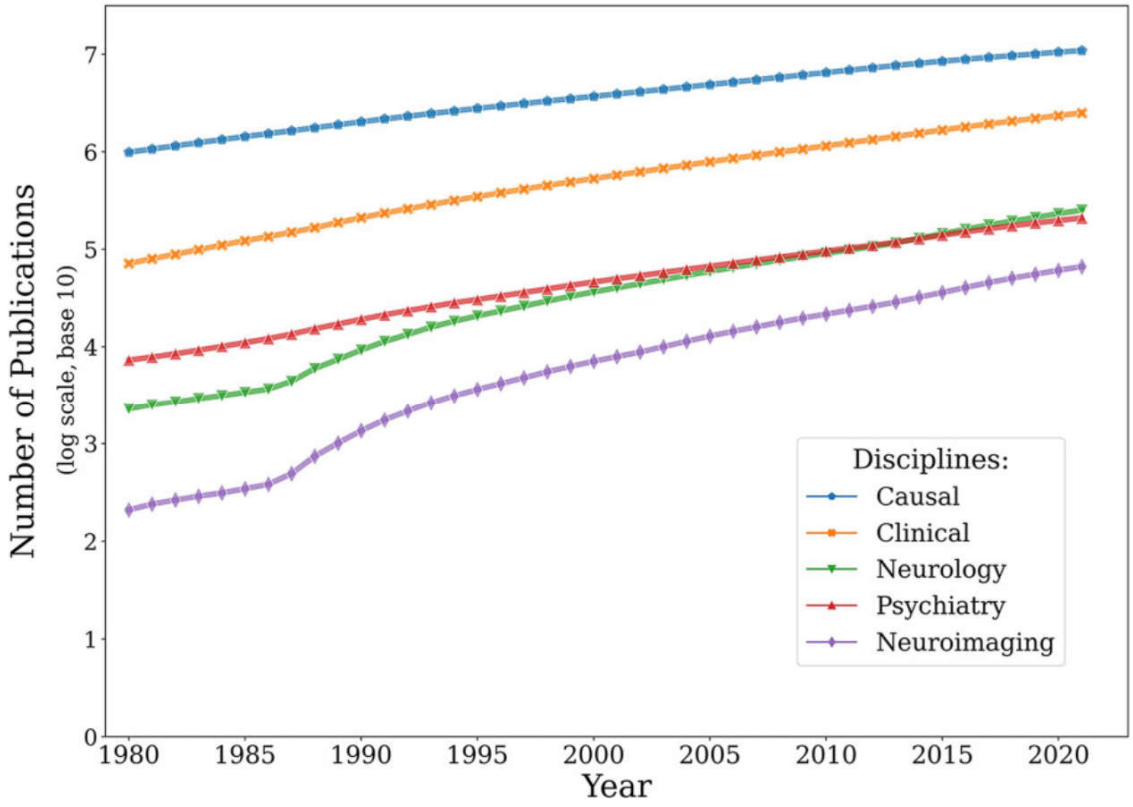
What you will learn in this tutorial?

1. Formulate your clinical question in a “causal way”

2. Basics for causal inference

3. How to estimate causal effect

The Rising Trend of Causal Related Research

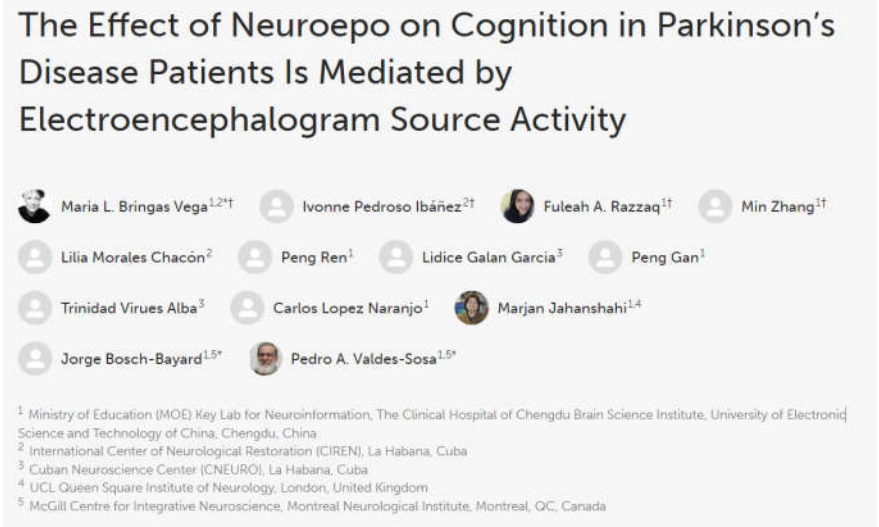


Qing Wang, et al. (2023) Psychoradiology.
https://github.com/Vincent-wq/causal_literature_trend

The previous NeuroEPO example

Instead of "**What are the effects of NeuroEPO for PD patients**?"
We are more interested in answering such type of questions like:
"What the effect will be if the PD patients (with some specific characteristics) take or not take NeuroEPO?"

Counterfactual Prediction


The Effect of Neuroepo on Cognition in Parkinson's Disease Patients Is Mediated by Electroencephalogram Source Activity

Maria L. Bringas Vega[1,2*†] Ivonne Pedroso Ibáñez[2†] Fuleah A. Razzaq[1†] Min Zhang[1†]
Lilia Morales Chacón[2] Peng Ren[1] Lidice Galan Garcia[3] Peng Gan[1]
Trinidad Virues Alba[3] Carlos Lopez Naranjo[1] Marjan Jahanshahi[1,4]
Jorge Bosch-Bayard[1,5*] Pedro A. Valdes-Sosa[1,5*]

[1] Ministry of Education (MOE) Key Lab for Neuroinformation, The Clinical Hospital of Chengdu Brain Science Institute, University of Electronic Science and Technology of China, Chengdu, China
[2] International Center of Neurological Restoration (CIREN), La Habana, Cuba
[3] Cuban Neuroscience Center (CNEURO), La Habana, Cuba
[4] UCL Queen Square Institute of Neurology, London, United Kingdom
[5] McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, Montreal, QC, Canada

Maria L Bringas Vega, et al (2022) Front. Neurosci.

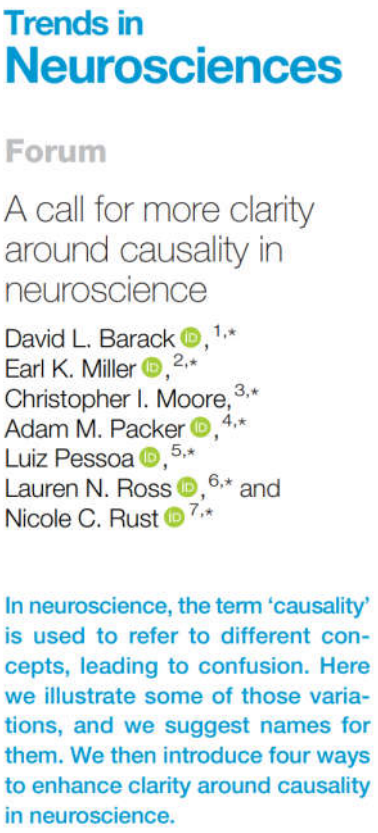
Basics of causal inference

- **Why causal inference is important?**
  - Most scientific research are causal in nature. However, traditional statistical analysis is to infer associations among variables, based on which we do a lot of predictions.

- **Why causal analysis is different?**
  - Causal analysis is about counterfactual predictions, predict what would have happened to the same units/subjects had they were exposed to a different (counterfactual) condition

- **In most cases, association does not imply causation!**

People are confusing about “causality”



Trends in
Neurosciences

Forum

A call for more clarity around causality in neuroscience

David L. Barack [1,*]
Earl K. Miller [2,*]
Christopher I. Moore, [3,*]
Adam M. Packer [4,*]
Luiz Pessoa [5,*]
Lauren N. Ross [6,*] and
Nicole C. Rust [7,*]

In neuroscience, the term ‘causality’ is used to refer to different concepts, leading to confusion. Here we illustrate some of those variations, and we suggest names for them. We then introduce four ways to enhance clarity around causality in neuroscience.

Barack, et al. (2022) Trends in Neuroscience.

- Causes are the events that produce other events ?
- Causes are the factors that events depend on?
- No causation without manipulation?

OXFORD

Psychoradiology, 2023, **3**, 1–4
DOI: 10.1093/psyrad/kkad007
Perspective

Claim causality with clarity

Qing Wang [1,*], Qiao Wang [2] and Ru-Yuan Zhang [1,3,*]

[1]Shanghai Mental Health Center, School of Medicine, Shanghai Jiao Tong University, 600 S. Wanping Road, Shanghai 200030, Shanghai, China
[2]Department of Biostatistics and Bioinformatics, Duke University, 2424 Erwin Road, Durham NC, 27705, USA
[3]Institute of Psychology and Behavioral Science, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China
*Correspondence: wangqing_vincent@smhc.org.cn; ruyuanzhang@sjtu.edu.cn

Qing Wang, et al. (2023) Psychoradiology.

Then, how?

- Causal assumptions and causal model (structure)

- The steps of causal inference

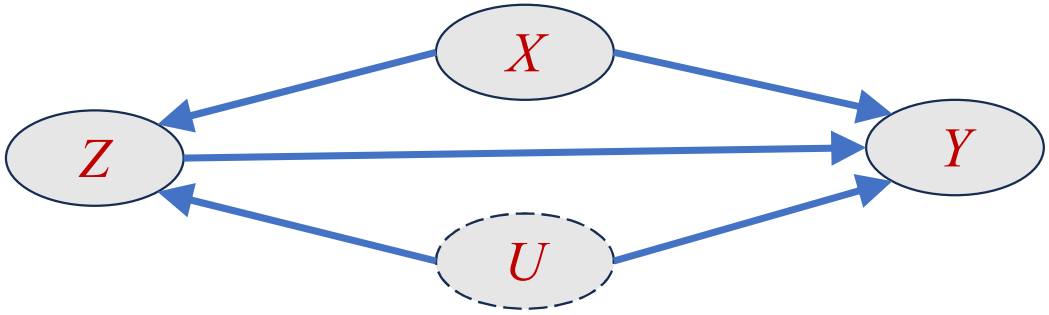
  1. Formulate your causal model (can be expressed by a graph)
  2. Specify all the assumptions under which, causation can be identified
  3. Carry out sensitivity analysis (with respect to causal assumptions and estimators)

Let us start with notations

- **Treatment** (e.g. intervention/exposure): Z (binary for tutorial purpose)
- **Outcome** (e.g. disease status (effect of treatment)): Y
- **Observed covariates or confounders**: X
- **Unobserved covariates or confounders**: U

Causal relationship

Confounding



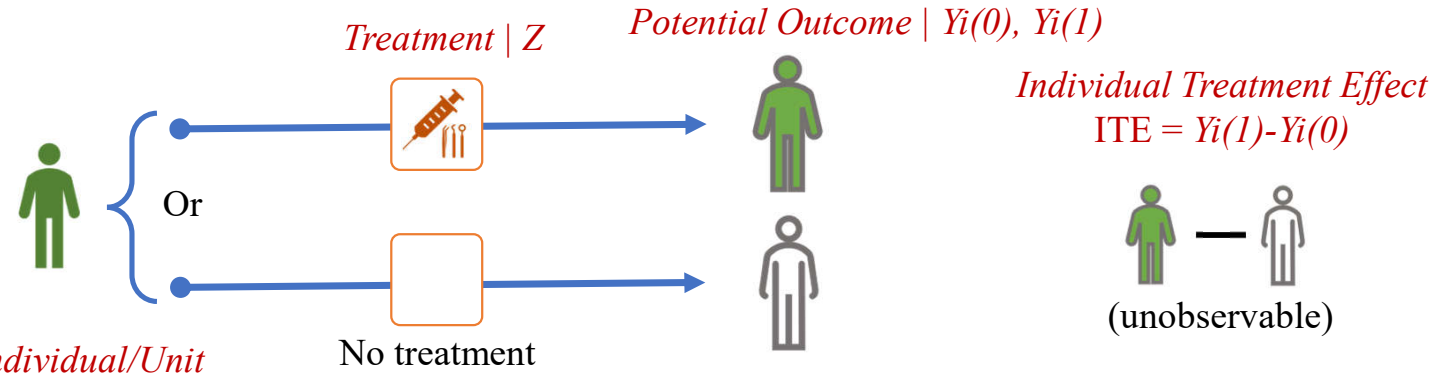
2 commonly used frameworks

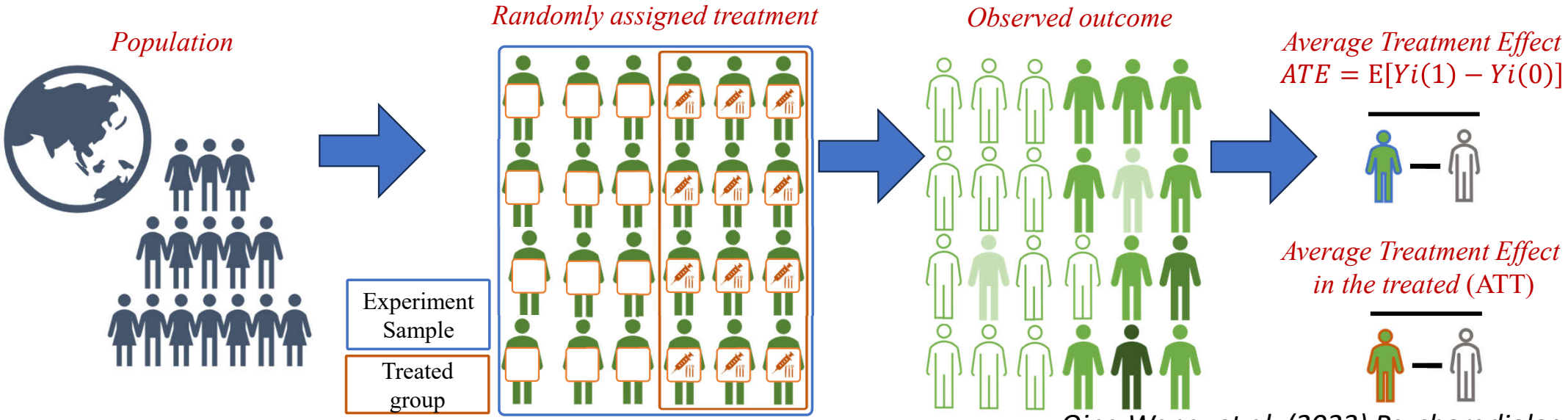
- The *potential outcome framework*, also know as the counterfactual framework, or Neyman-Rubin Causal Model (Neyman, 1923; Rubin, 1974; Imbens and Rubin, 2015; Hernan and Robins, 2020)

- The *Structural Causal Model* or causal diagram framework (Pearl, 2009)

- Mathematically they are connected (Richardson and Robins, 2013), but with different established goals, tools and applications.

Potential Outcome Framework: Basic setup





Qing Wang, et al. (2023) Psychoradiology.

The fundamental challenge for causal inference

- The fundamental problem of causal inference: **We can observe at most one of the potential outcomes for each unit**, the other(s) are missing/counterfactual

- Causal inference under the potential outcome framework is essentially a **missing data problem**

- To identify causal effects from observational data, under any mathematical framework, one must make assumptions (structural or/and stochastic)

Holland, (1986) JASA

Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980)

- What is SUTVA?
  - No interference between unites
  - No different versions of a treatment (also known as Consistency, e.g. doses etc…)

- What does this assumption offer?
  - For binary treatment, we only have 2 potential outcomes: $\mathrm{Yi}(0)$, $\mathrm{Yi}(1)$
  - If $Z_i = 1$ then $Y_i = Y_i(1)$
    If $Z_i = 0$ then $Y_i = Y_i(0)$
    Equivalently: $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$

SUTVA connects the intervention we see (Z) with the causal intervention of interest (z)

Causal estimands (causal effect of interest)

- **Conditional average treatment effect (CATE),** also known as individual treatment effect (ITE) conditional on a covariate value

$$\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X = x]$$

- **Average treatment effect (ATE):**

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}_x[\tau(x)].$$

- **Average treatment effect for the treated (ATT):**

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)|Z_i = 1].$$

- **Average treatment effect for the control (ATC):**

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)|Z_i = 0].$$

It can be ratios as well

$$r_i = Y_i(1)/Y_i(0)$$

Potential outcomes

- Causal effects are defined by potential outcomes, not model parameters
- We introduce this “hypothetical” potential outcome for
  - Disentangle intervention from quantities of interest
  - Randomness of the assignment of treatment breaks the confounding effect from ancestors of both Z and Y
  - Analysis of experiments is driven by design
- We still need further assumptions to identify the causal estimands

Unconfounded assignment assumption

- Unconfounded Assignment: an assignment is unconfounded if the assignment mechanism does not depend on the potential outcomes (and only on the pre-treatment covariates)

$$p(Z_i = 1|X_i, Y_i(0), Y_i(1)) = p(Z_i = 1|X_i)$$

- Yes, this is the Propensity Score! (Rosenbaum and Rubin, 1983)
- This is also why RCT is the "Golden Standard" for estimating causal effect
- Also known as Ignorable Assignment or ignorability

Be clear about your study type

- **Randomized Experiments**: the assignment mechanism is, known, controlled, and random (stronger than unconfounded)

- **Observational studies**: the assignment mechanism is unknown and uncontrolled, but often assumed to be unconfounded conditional on observed covariates or unobserved quantities
  - Cross-sectional data: treatment at one time point
  - Longitudinal data: treatment at multiple time points (sequentially ignorable)
  - Panel data: treatment at one time point, comparative case studies (often different assumptions than unconfoundedness)

Methods and modes of inference

- Two overarching methods:
  - Imputation: impute the missing potential outcomes (model-base or matching-based)
  - Weighting: weight (often function of the propensity scores) the observed data to represent a target population

- Three modes of inference
  - Frequentist: imputation, weighting, motivated by consistency, asymptotic normality, (semiparametric) efficiency, etc.
  - Bayesian: modeling and imputing missing potential outcomes based on their posterior distributions
  - Fisherian randomization: combine randomization tests with Bayesian methods, unique to randomized experiments

Causal inference V.S. Missing data

- Under PO framework, **causal inference is a missing data problem**

- A broad parallel between the classification of assignment mechanisms in causal inference and the classification of missing data mechanisms (Ding and Li, 2018, Stat Sci)
  - Completely randomized experiments ↔missing completely at random (MCAR)
  - Observational studies with unmeasured confounding ↔ missing not at random (MNAR)
  - ignorable assignment mechanisms MAR ↔ missing at random(MAR)

Other sources of bias

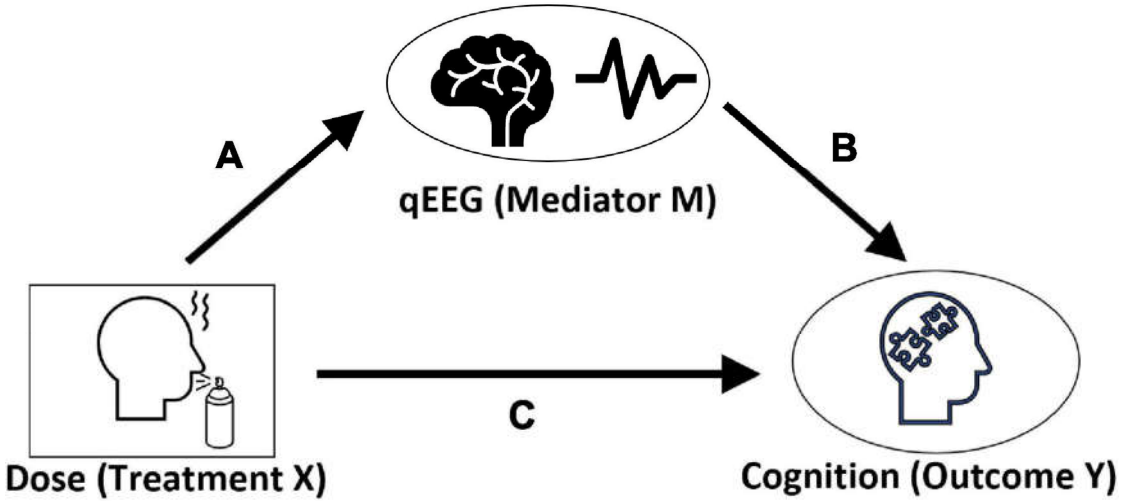
- **Selection Bias**: biases that arise from the procedure by which individuals are selected into the analysis (Hernan and Robins, 2020), notice the definition of “target population” (Sample ATE etc.)

- **Measurement Bias**: comes from the strong assumption of assumption that all variables were perfectly measured, which may be unrealistic, think about neuroimaging, EEG and qEEG, etc.

What is mediation effect?

Structural Equation Modeling (SEM)

- Mediation Effect
- Direct Effect
- Total Effect



Mediator model:

qEEG~1+Dose+severity+education+age.trial+random(Subjects)

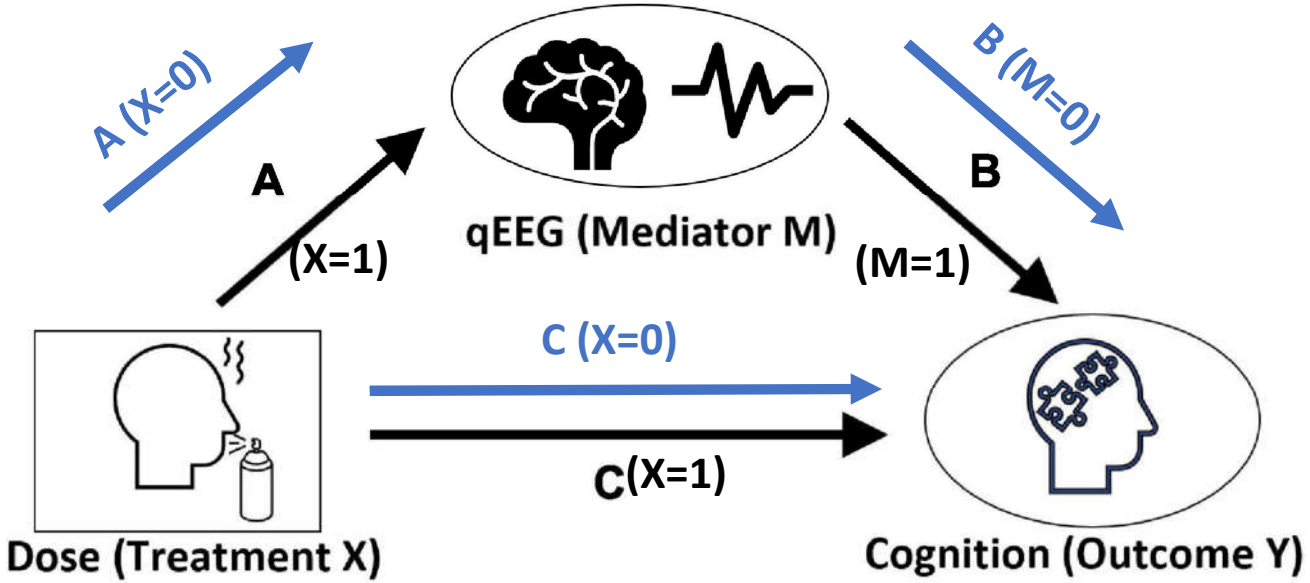
Outcome model:

Cog~1+qEEG+Dose+severity+education+age.trial+random(Subjects)

Maria L Bringas Vega, et al (2022) Front. Neurosci.

What is causal mediation effect?

- Average Causal Mediation Effect (ACME)
- Average Direct Effect (ADE)
- Sensitivity Analysis/Refutation tests



Qin Xu, et al. (2024) Beh. Res. Meth.

Congrats!

You have already learned the basic concepts of causal inference (from a statistician's perspective), let us advance with an example!

Tools for causal inference, there are lots more…

- If you like R more:
  - PSweight: https://cran.r-project.org/web/packages/PSweight/index.html,
  - Mediation: (https://cran.r-project.org/web/packages/mediation/index.html,
  - etc.

- If you like python more:
  - DoWhy: https://py-why.github.io/dowhy
  - EconML: https://github.com/microsoft/EconML
  - etc.
- There are so many tools and more new tools are still in development…

Example settings

Figure 1

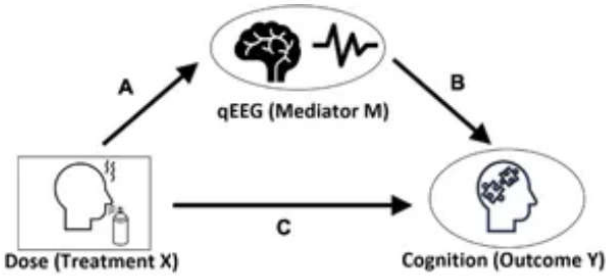


Figure 1. A mediation model. Effect of dosage on Cognition *via* qEEG. All three are repeated measures at two-time points. Ovals are latent variables, whereas rectangle shows observed variables. In this causal diagram, path c is the "direct effect of the Dose on Cognition." Path following links a and b represent the "mediation effect" we wish to test.

Notice: this example is based on simulation only for tutorial purposes, it is different from the original paper which use the data collected from and RCT.

Codes can be found here: https://github.com/Vincent-wq/causal_course_eeg

Step1. Specify model with Graph Modeling Language

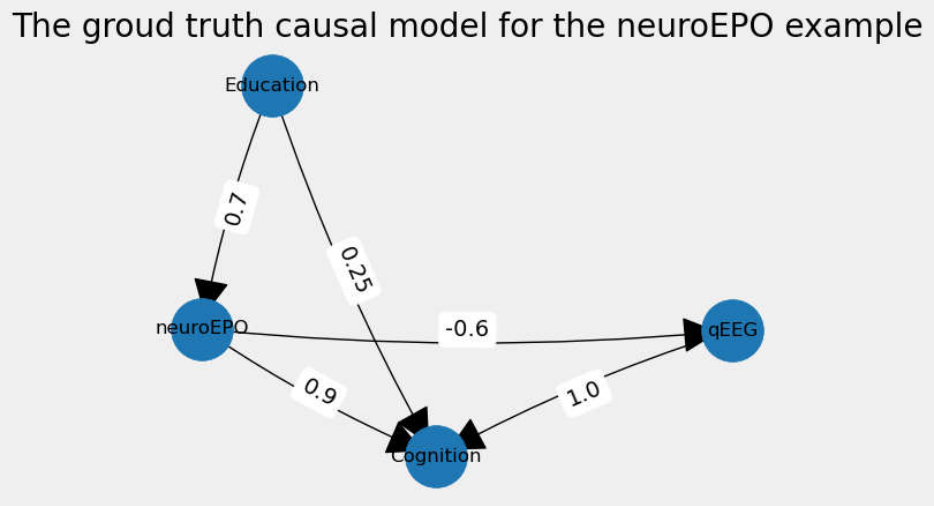
```
# Create the graph describing the causal structure
gt_graph = """
graph [
    directed 1

    node [
        id "Z"
        label "neuroEPO"
    ]
    node [
        id "X"
        label "education"
    ]
    node [
        id "M"
        label "qEEG"
    ]
    node [
        id "Y"
        label "cognition"
    ]
    edge [
        source "X"
        target "Z"
    ]
    node [
        id "Y"
        label "cognition"
    ]
    edge [
        source "X"
        target "Z"
    ]
    edge [
        source "X"
        target "Y"
        weight  0.6
    ]
    edge [
        source "Z"
        target "Y"
    ]
    edge [
        source "Z"
        target "M"
    ]

    edge [
        source "M"
        target "Y"
    ]
]
"""
```

nodes

edges



Step2. Define the model and specify the estimand

```python
# define the model
model = CausalModel(
    data=df,
    treatment='neuroEPO',
    outcome='Cognition',
    common_causes = 'Education',
    graph=neuroEPO_graph
)
print("The graphical model specified:")
```

Step 2: Identify the estimand

```python
estimand = model.identify_effect()
print(estimand)
```

[17] ✓ 0.1s Python

```
Estimand type: EstimandType.NONPARAMETRIC_ATE

### Estimand : 1
Estimand name: backdoor
No such variable(s) found!

### Estimand : 2
Estimand name: iv
No such variable(s) found!

### Estimand : 3
Estimand name: frontdoor
Estimand expression:
  ⎡  d       d       ⎤
E⎢─────(Y)⋅─────([Z])⎥
  ⎣d[Z]    d[X]      ⎦
Estimand assumption 1, Full-mediation: Z intercepts (blocks) all directed paths from X to Y.
Estimand assumption 2, First-stage-unconfoundedness: If U→{X} and U→{Z} then P(Z|X,U) = P(Z|X)
Estimand assumption 3, Second-stage-unconfoundedness: If U→{Z} and U→Y then P(Y|Z, X, U) = P(Y|Z, X)
```

Step3. Sensitivity analysis / refutation test

```
print(random_cause)
```

```
Refute: Add a random common cause
Estimated effect:0.6528130665938017
New effect:0.6737806276459242
p value:0.30000000000000004
```

```python
random_cause = model.refute_estimate(
    estimand=estimand,
    estimate=estimate,
    method_name='random_common_cause'
)
```

Python

```
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ),
```

What our estimands will behave if some of the assumptions are violated?

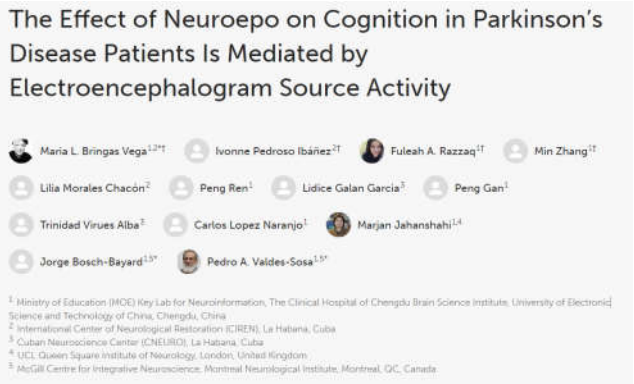
Take home message

- Most scientific research are causal in nature, thus causal reasoning is inevitable
- EEG and qEEG are important, and of course not ideal observation of the “brain functioning”, their roles in clinical neuroscience research still need further exploration
- Tools and software for causal analysis are already there and evolving, no matter which programming language you prefer
- Open access data, pipelines and platforms allow us to improve the generalizability and interpretability of our research and do take this advantage!

Oh! Homework!

- Clinical Application Registry:
  - https://classic.clinicaltrials.gov/ct2/show/NCT04110678

- EEG data are freely available on OpenNEURO
  - Neuroepo multisession: https://openneuro.org/datasets/ds003194/versions/1.0.0
  - Placebo Neuroepo multisession: https://openneuro.org/datasets/ds003195/versions/1.0.0


The Effect of Neuroepo on Cognition in Parkinson's Disease Patients Is Mediated by Electroencephalogram Source Activity

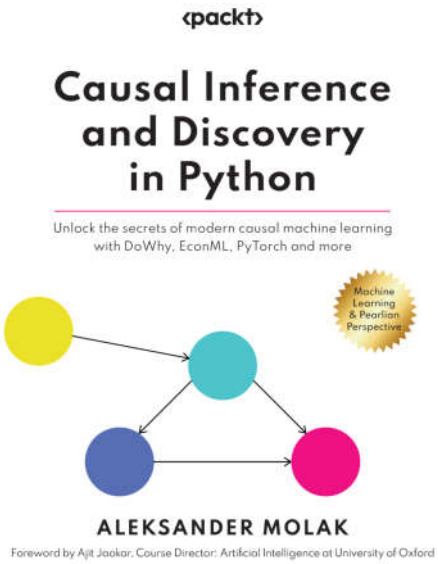
Maria L. Bringas Vega, Ivonne Pedroso Ibáñez, Fuleah A. Razzaq, Min Zhang, Lilia Morales Chacón, Peng Ren, Lidice Galan Garcia, Peng Gan, Trinidad Virues Alba, Carlos Lopez Naranjo, Marjan Jahanshahi, Jorge Bosch-Bayard, Pedro A. Valdes-Sosa

Maria L Bringas Vega, et al (2022) Front. Neurosci.

Some other interesting tutorials

- Causal Inference course (Fan Li, Duke): https://www2.stat.duke.edu/~fl35/CausalInferenceClass.html

- McGill QLSC612 : Foundations of Neuro Data Science, https://neurodatascience.github.io/QLS612-Overview/



Thank you!

Enjoy this course and the annual meeting!

OHBM 2024

JUNE 23~27, SEOUL, KOREA