Genes, life-course and heart failure – intro to PhD project

Anders Munch

September 30, 2020

Introduction to PhD project

- The problem
- The data
- Some thoughts/ideas so far

People involved

- Thomas Gerds and Claus Ekstrøm from here
- Christian Torp-Pedersen, Professor of Cardiology and Senior Consultant, Gentofte Hospital and University of Copenhagen
- Charlotte Andersson, MD PhD, Herlev Hospital and Boston Medical Center

Overall motivation / "hypothesis"

- Better understanding of the causes of HF. In particular a focus on the genetic effects.

Overall motivation / "hypothesis"

- Better understanding of the causes of HF. In particular a focus on the genetic effects.
- Heterogeneous disease / syndrome.

Overall motivation / "hypothesis"

- Better understanding of the causes of HF. In particular a focus on the genetic effects.
- Heterogeneous disease / syndrome.
- Better HF sub-diagnoses or disease pattern $\rightarrow$ identify genetic effects.

Overall motivation / "hypothesis"

- Better understanding of the causes of HF. In particular a focus on the genetic effects.
- Heterogeneous disease / syndrome.
- Better HF sub-diagnoses or disease pattern $\rightarrow$ identify genetic effects.
- Use machine learning for this ...

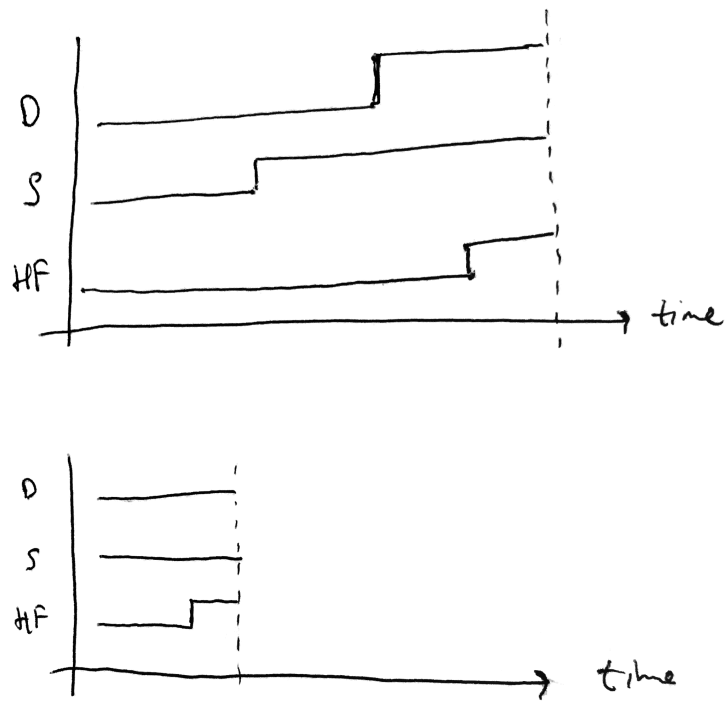
Data

Coronary arteries examiniation: $\sim 6,000$ individuals + DST data. Weird selection criteria, non-representative population.

Blood donors: $\sim 300,000$ individuals in "computer room". Very healthy individuals.

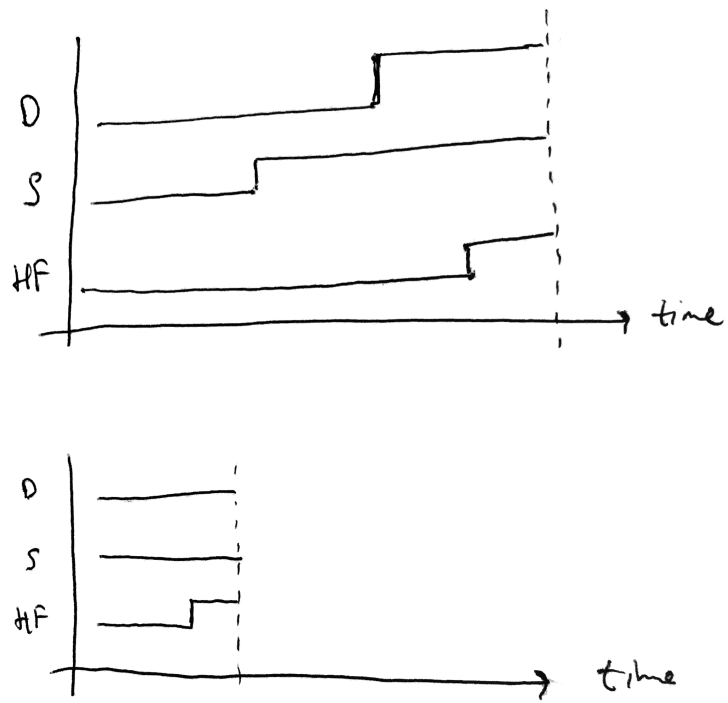
Framingham: Participants from a small city. Detailed medical information measured repeatedly over many years and generations. Genetic data for later generations.

Initial thoughts



Initial thoughts

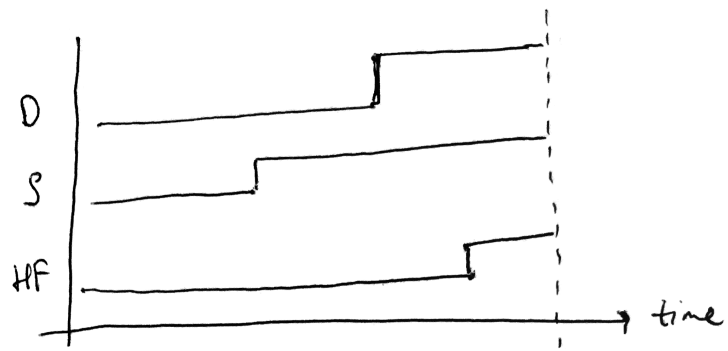
Cluster disease patterns



Initial thoughts

Cluster disease patterns

Mediation



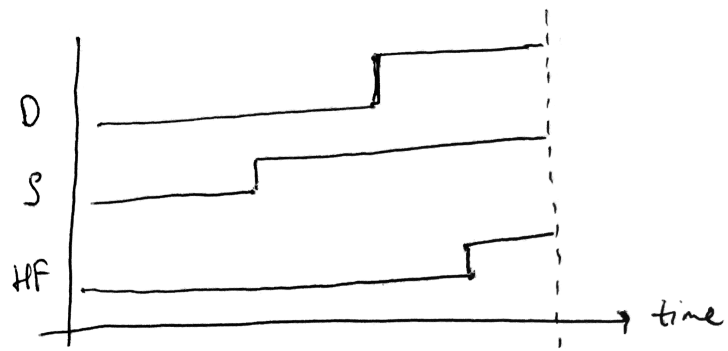
Initial thoughts

Cluster disease patterns

Mediation

↓

Local independence?



Formalization of the problem

One rough initial formalization

We observe a multivariate process $X = (G, D, Y)$,

$$Y(t), \quad D(t), \quad G(t) = G, \quad t \in [0, T],$$

where $Y \in \{0, 1\}$ denotes if heart failure has occurred, D denotes status of some disease (for instance diabetes or not at time t), G is constant and contains genetic information, and T is time of death.

Formalization of the problem

One rough initial formalization

We observe a multivariate process $X = (G, D, Y)$,

$$Y(t), \quad D(t), \quad G(t) = G, \quad t \in [0, T],$$

where $Y \in \{0, 1\}$ denotes if heart failure has occurred, D denotes status of some disease (for instance diabetes or not at time t), G is constant and contains genetic information, and T is time of death.

Later

- much more complicated D
- very high-dimensional G

For now, focus on the time dynamic issues so we assume both are one-dimensional.

Informal questions we might want to ask

Informal questions we might want to ask

- Are there some particular disease patterns for which the genetic disposition is particularly important for the risk of HF?

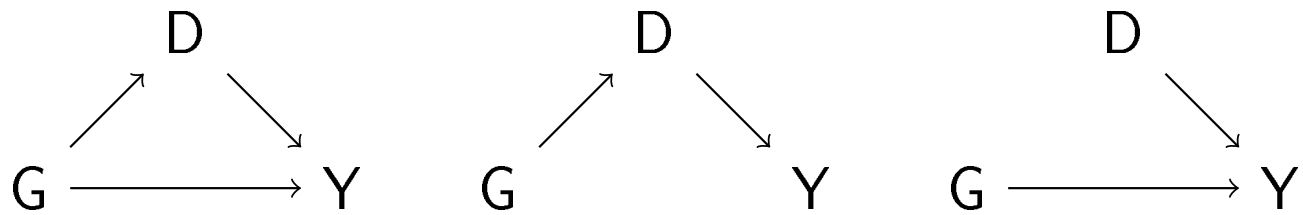
Informal questions we might want to ask

- Are there some particular disease patterns for which the genetic disposition is particularly important for the risk of HF?
- How does the genetic mechanisms work? Direct influence on the risk of HF or indirect influence through increased risk of precursors for HF?

Informal questions we might want to ask

- Are there some particular disease patterns for which the genetic disposition is particularly important for the risk of HF?
- How does the genetic mechanisms work? Direct influence on the risk of HF or indirect influence through increased risk of precursors for HF?

Doubtful illustrations



What does the above mean when D and Y are processes?

Local independence

Local independence

Definition (Informal)

For a multivariate stochastic X

$$X(t) = (X^1(t), X^2(t), \ldots, X^k(t)), \quad t \in [0, T],$$

with $V := \{1, \ldots, k\}$, we say that for $A, B, C \subset V$, X^B is *locally independent* of X^A given X^C if

$$X^B(t) \perp\!\!\!\perp X^A([0, t)) \mid X^C([0, t)), \quad \forall t \in [0, T].$$

This is written as $A \not\to B \mid C$.

Local independence

Definition (Informal)

For a multivariate stochastic X

$$X(t) = (X^1(t), X^2(t), \ldots, X^k(t)), \quad t \in [0, T],$$

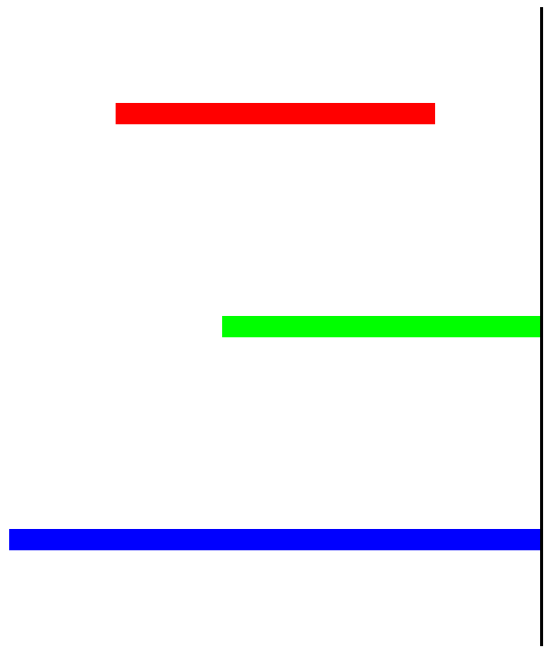
with $V := \{1, \ldots, k\}$, we say that for $A, B, C \subset V$, X^B is *locally independent* of X^A given X^C if

$$X^B(t) \perp\!\!\!\perp X^A([0, t)) \mid X^C([0, t)), \quad \forall t \in [0, T].$$

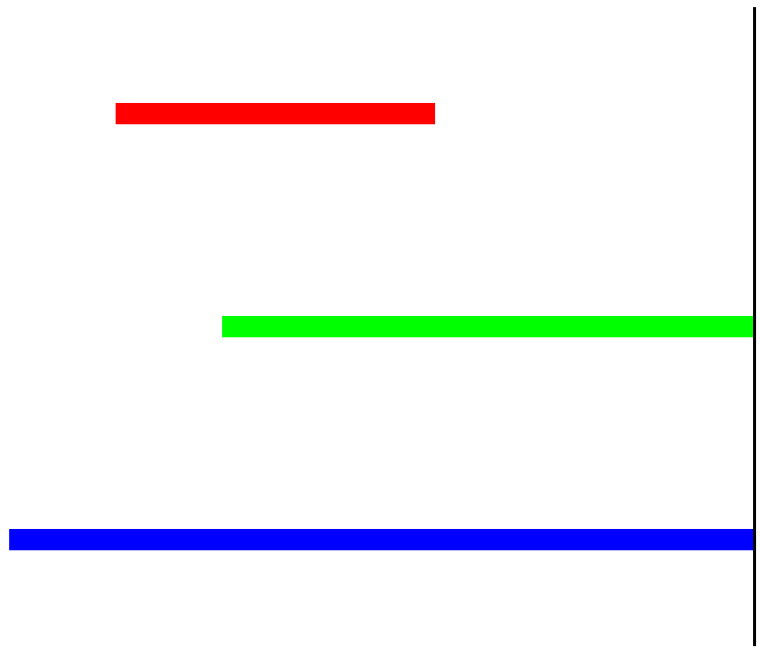
This is written as $A \not\to B \mid C$.

"In words, the process X^B is locally independent of X^A given X^C if, for each time point, the past up until time t of X^C gives us the same *predictable* information about $\mathbb{E}[X^\beta(t) \mid \mathcal{F}_t^{A \cup C}]$ as the past of $X^{A \cup C}$ until time t." [Mogensen et al., 2020]

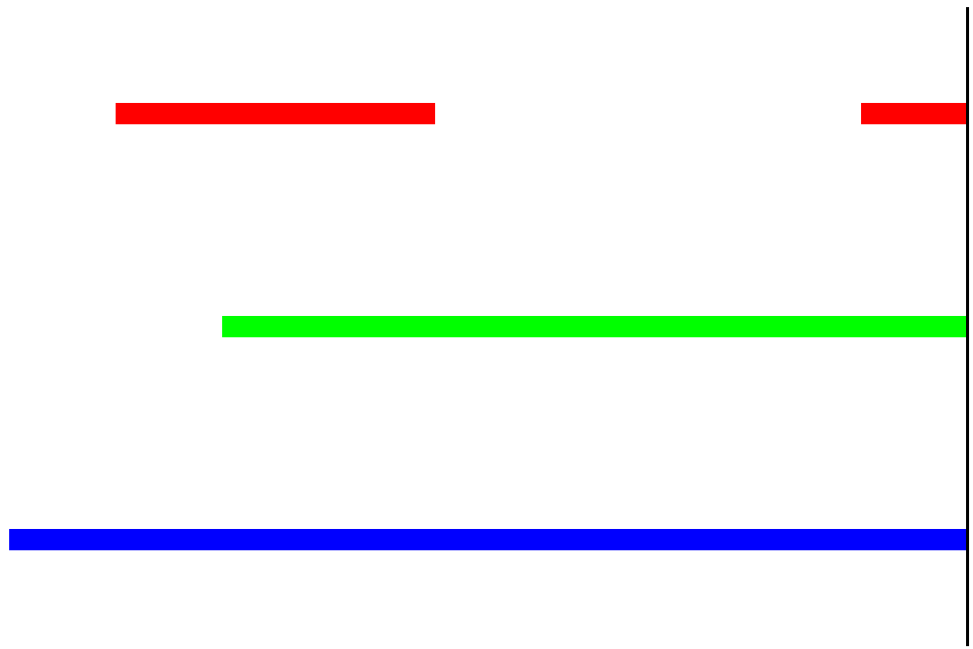
Local independence – visualization



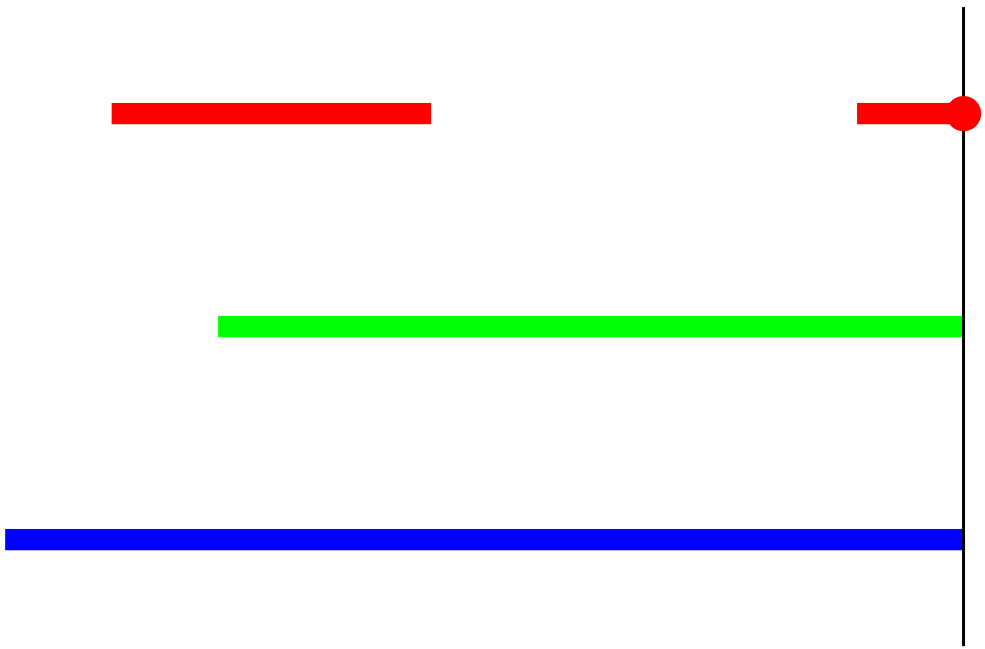
Local independence – visualization



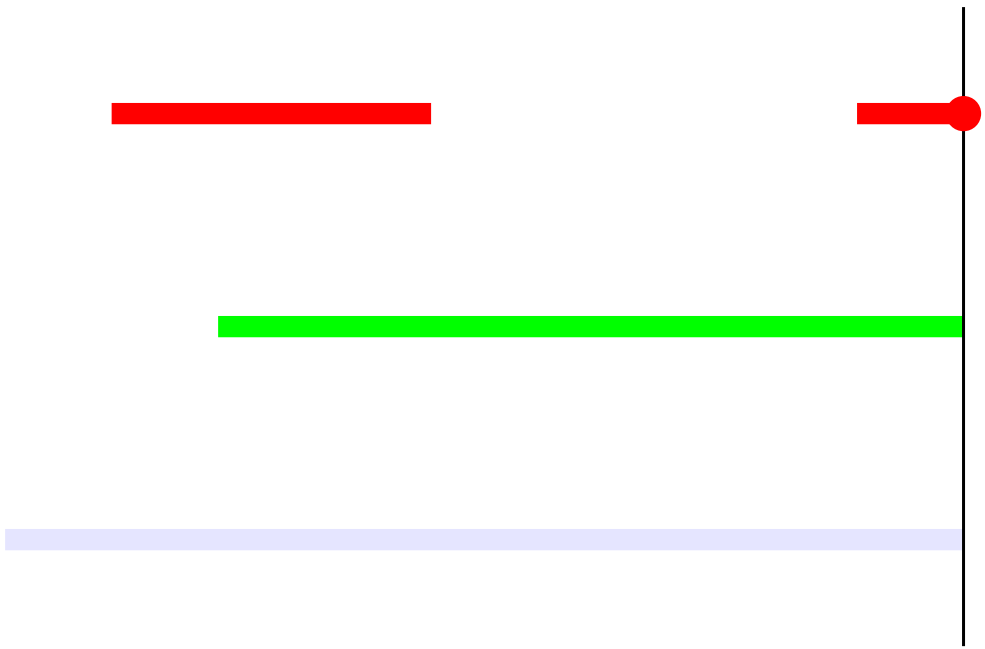
Local independence – visualization



Local independence – visualization



Local independence – visualization



Local independence – history

Introduced by Schweder [1970] and elaborated and applied in Aalen [1987] and Aalen et al. [1980].

Local independence – history

Introduced by Schweder [1970] and elaborated and applied in Aalen [1987] and Aalen et al. [1980].

Relation to graphical models considered by Didelez [2008] and Mogensen et al. [2020].

Local independence – history

Introduced by Schweder [1970] and elaborated and applied in Aalen [1987] and Aalen et al. [1980].

Relation to graphical models considered by Didelez [2008] and Mogensen et al. [2020].

Aalen et al. [2016] and Aalen et al. [2012] discuss and give examples as to why local independence might be better suited for modeling (causal) dependencies in a dynamical system.

Local independence – history

Introduced by Schweder [1970] and elaborated and applied in Aalen [1987] and Aalen et al. [1980].

Relation to graphical models considered by Didelez [2008] and Mogensen et al. [2020].

Aalen et al. [2016] and Aalen et al. [2012] discuss and give examples as to why local independence might be better suited for modeling (causal) dependencies in a dynamical system.

> *"We suggest that when people attempt to draw causal diagrams it is often most natural to think of the nodes as processes and use local independence" [Aalen et al., 2016, p.2300].*

How might local independence be relevant in our setting?

How might local independence be relevant in our setting?

1. Model of temporal dependence

How might local independence be relevant in our setting?

1. Model of temporal dependence
2. An approach to mediation in a dynamical system

How might local independence be relevant in our setting?

1. Model of temporal dependence
2. An approach to mediation in a dynamical system
3. “Dynamic” point of view

1. Clustering disease patterns

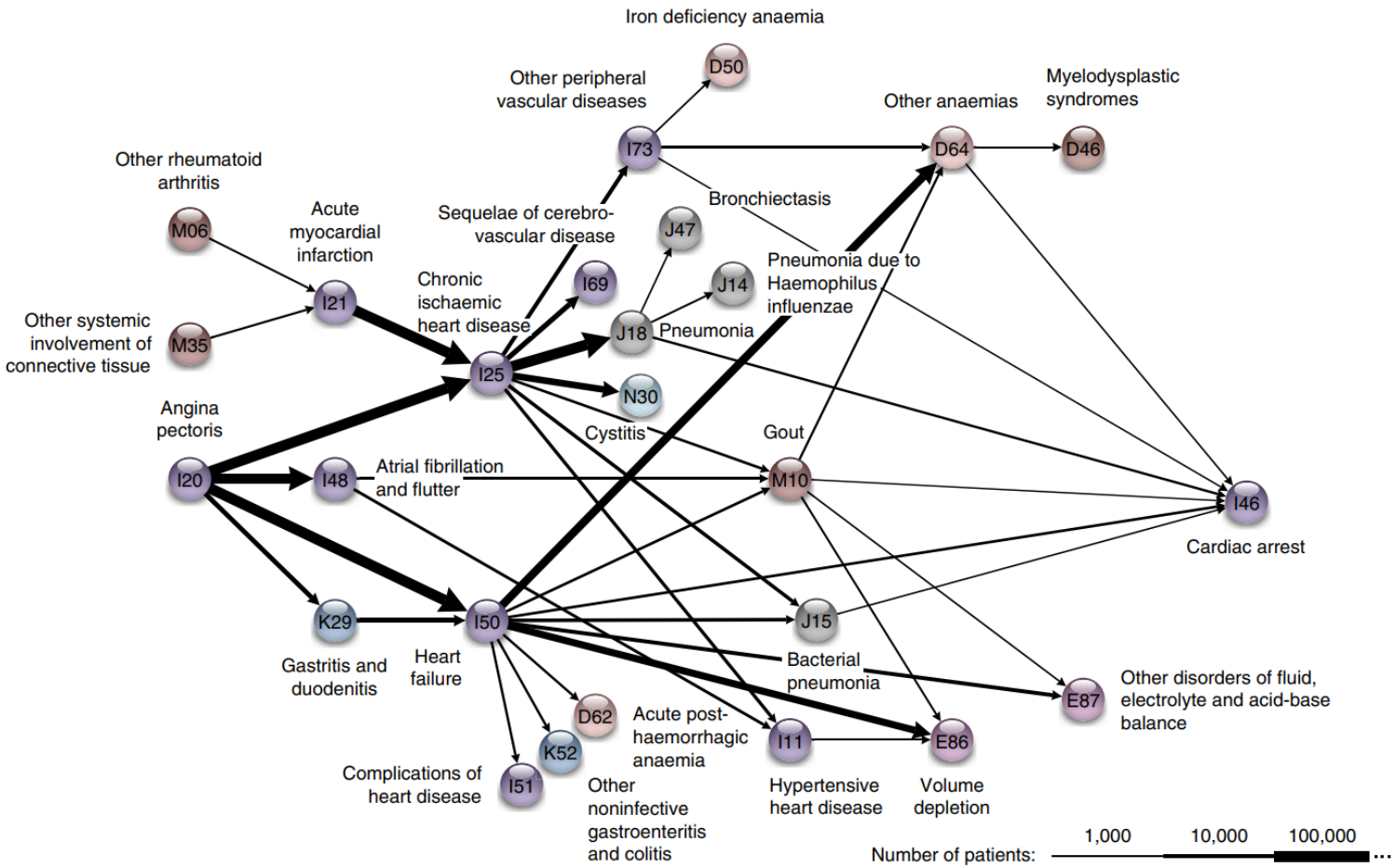
Brunak Group

Jensen et al. [2014] used Danish registries data to identify and cluster disease progressions.

1. Clustering disease patterns

Brunak Group

Jensen et al. [2014] used Danish registries data to identify and cluster disease progressions.



1. Clustering disease patterns – use local dependence?

Brunak Group's approach to clustering

- Matching every exposed patient to a non-exposed group with similar age and sex.
- Test for association between diagnoses occurring within 4 years, and then test for "temporal direction".
- Various ad hoc choices.
- Censoring and death not taken into account.
- Interpretation unclear.

1. Clustering disease patterns – use local dependence?

Brunak Group's approach to clustering

- Matching every exposed patient to a non-exposed group with similar age and sex.
- Test for association between diagnoses occurring within 4 years, and then test for "temporal direction".
- Various ad hoc choices.
- Censoring and death not taken into account.
- Interpretation unclear.

Try to use local dependence?

- Better way to model "temporal association"? Nicer interpretation?
- Easier to handle stopped processes – take death and censoring into account.
- Perhaps fewer arbitrary choices.

2. Mediation analysis in time

Mediation analysis in time seems to be complicated.

2. Mediation analysis in time

Mediation analysis in time seems to be complicated.

Example:

> *"To [...] define the relevant direct and indirect effects in the survival context, we will have to allow for hypothetical interventions on survival. [...] The difficulty that otherwise arises is that for those who do not survive the subsequent values of the mediator are always undefined. [...] While the mathematical development is precise, the interpretation of what such an intervention on survival means is ambiguous." [Lin et al., 2017, p.4154]*

2. Mediation analysis in time

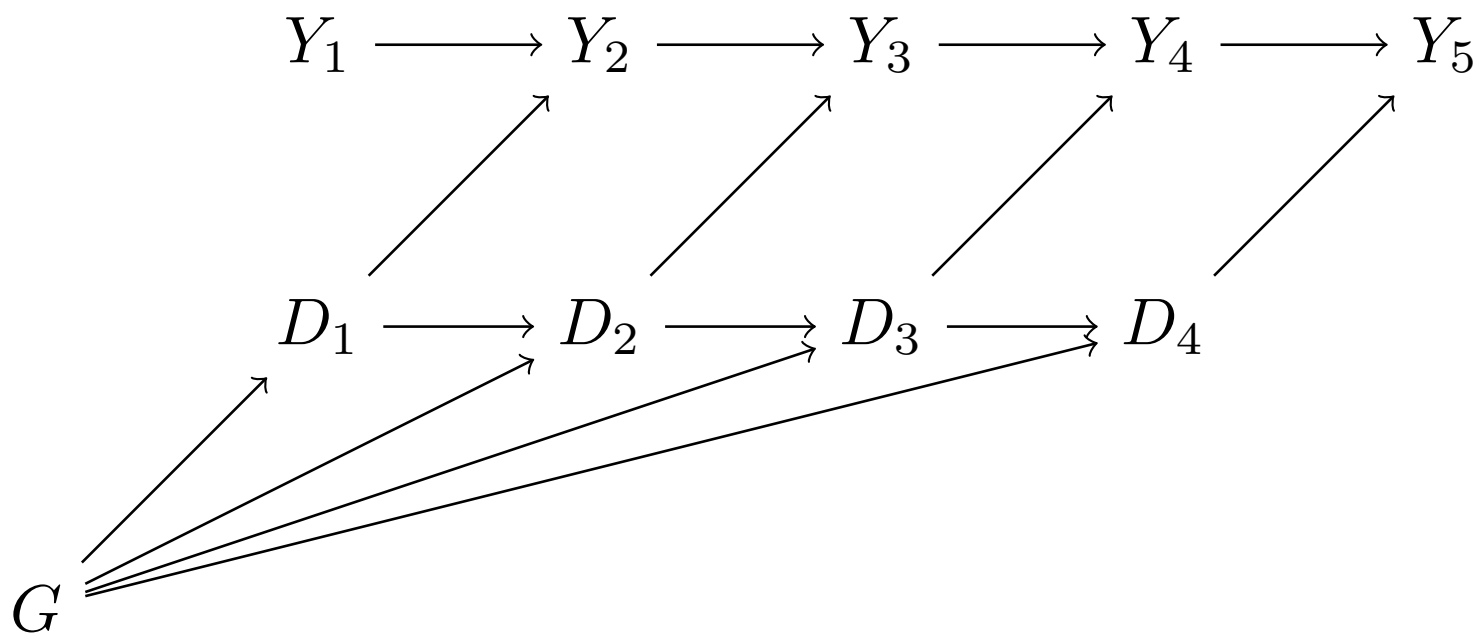
Mediation analysis in time seems to be complicated.

Example:

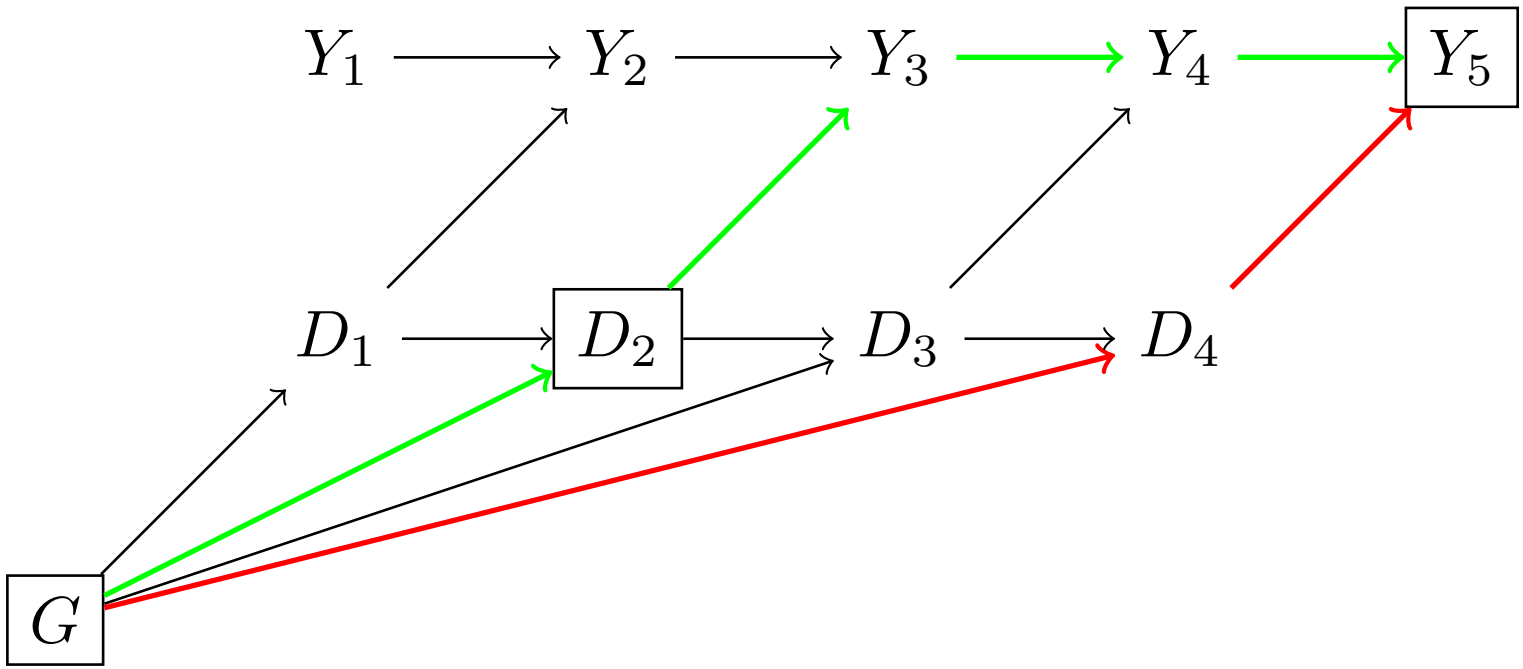
> *"To [...] define the relevant direct and indirect effects in the survival context, we will have to allow for hypothetical interventions on survival. [...] The difficulty that otherwise arises is that for those who do not survive the subsequent values of the mediator are always undefined. [...] While the mathematical development is precise, the interpretation of what such an intervention on survival means is ambiguous." [Lin et al., 2017, p.4154]*

In addition, causal mechanisms can be "smeared out" or "distorted" by discretization [Aalen et al., 2016].

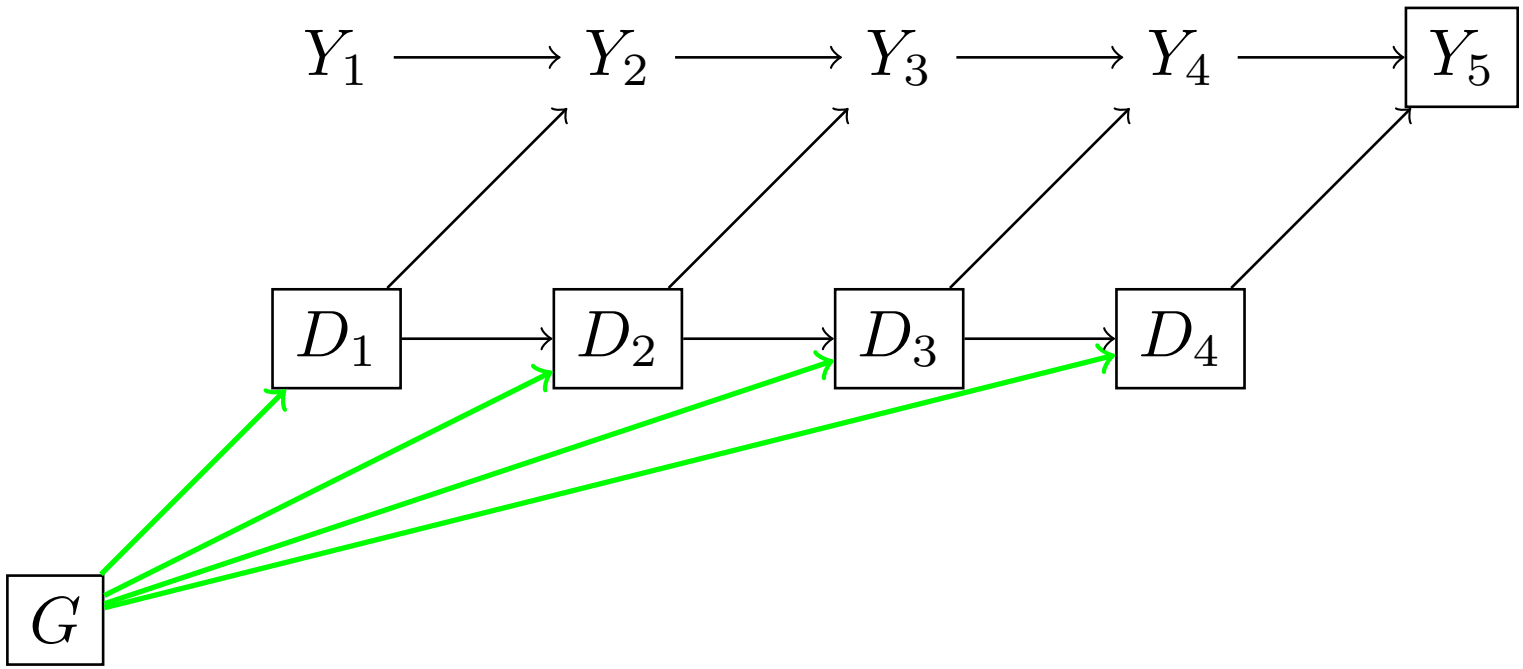
2. Mediation in time – visualization



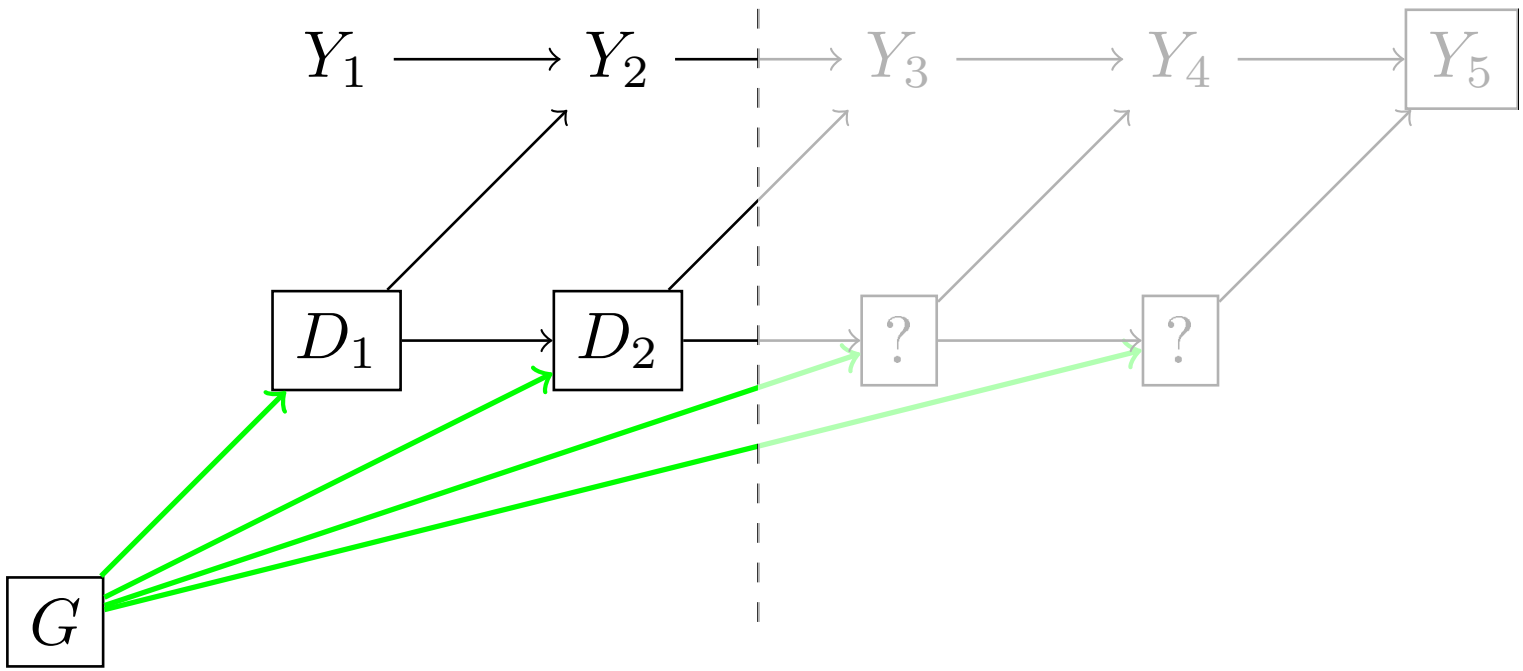
2. Mediation in time – visualization



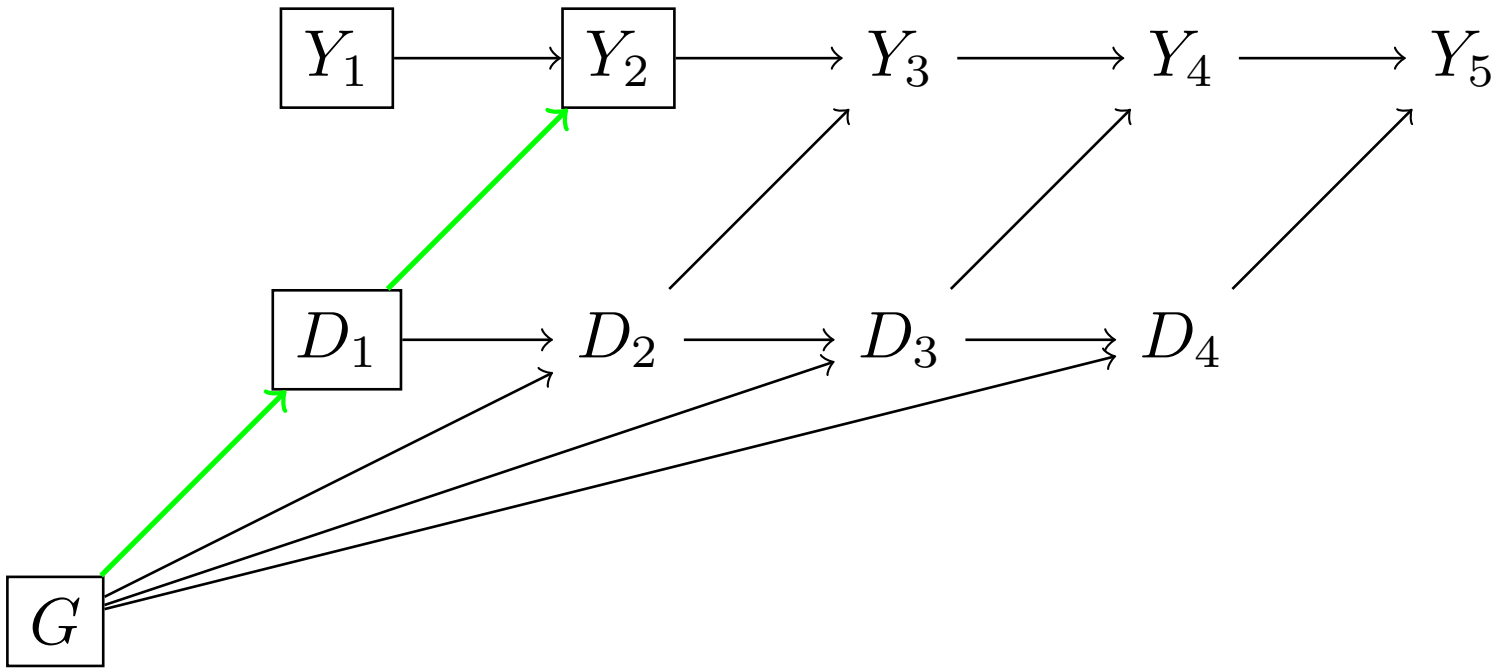
2. Mediation in time – visualization



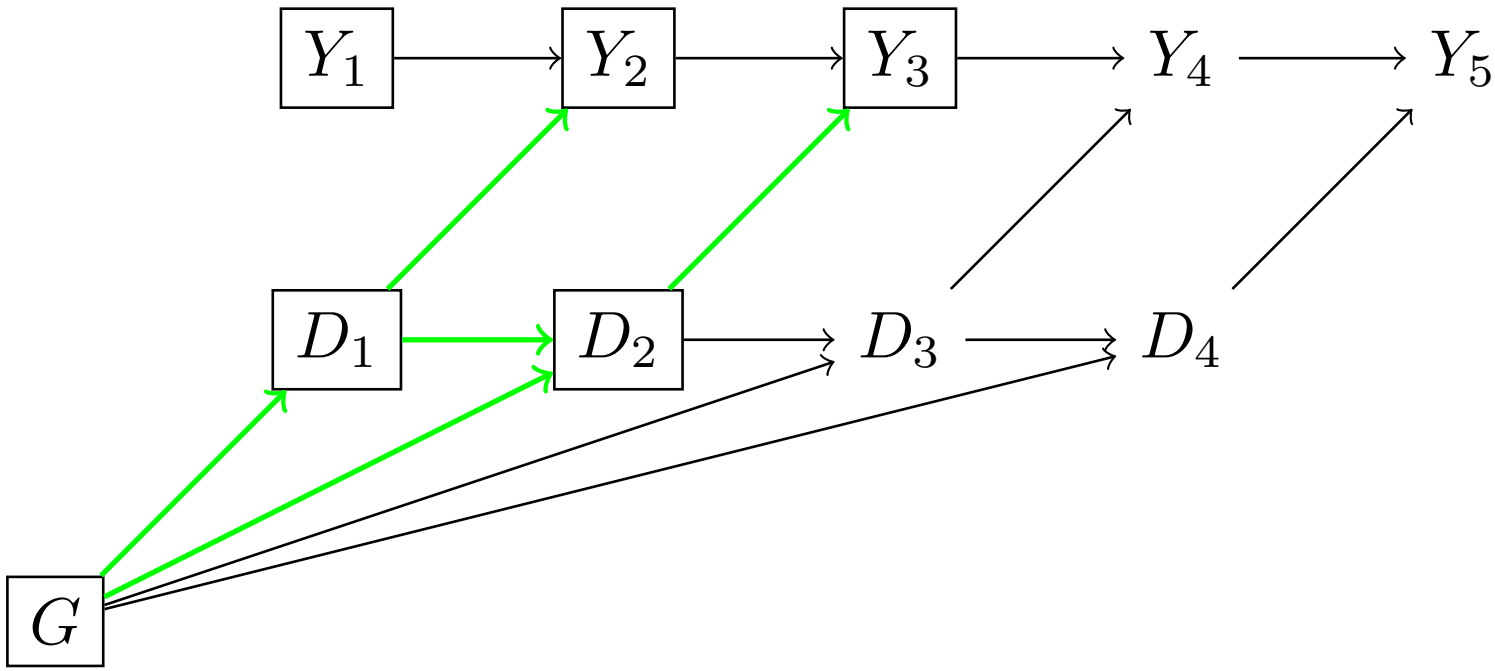
2. Mediation in time – visualization



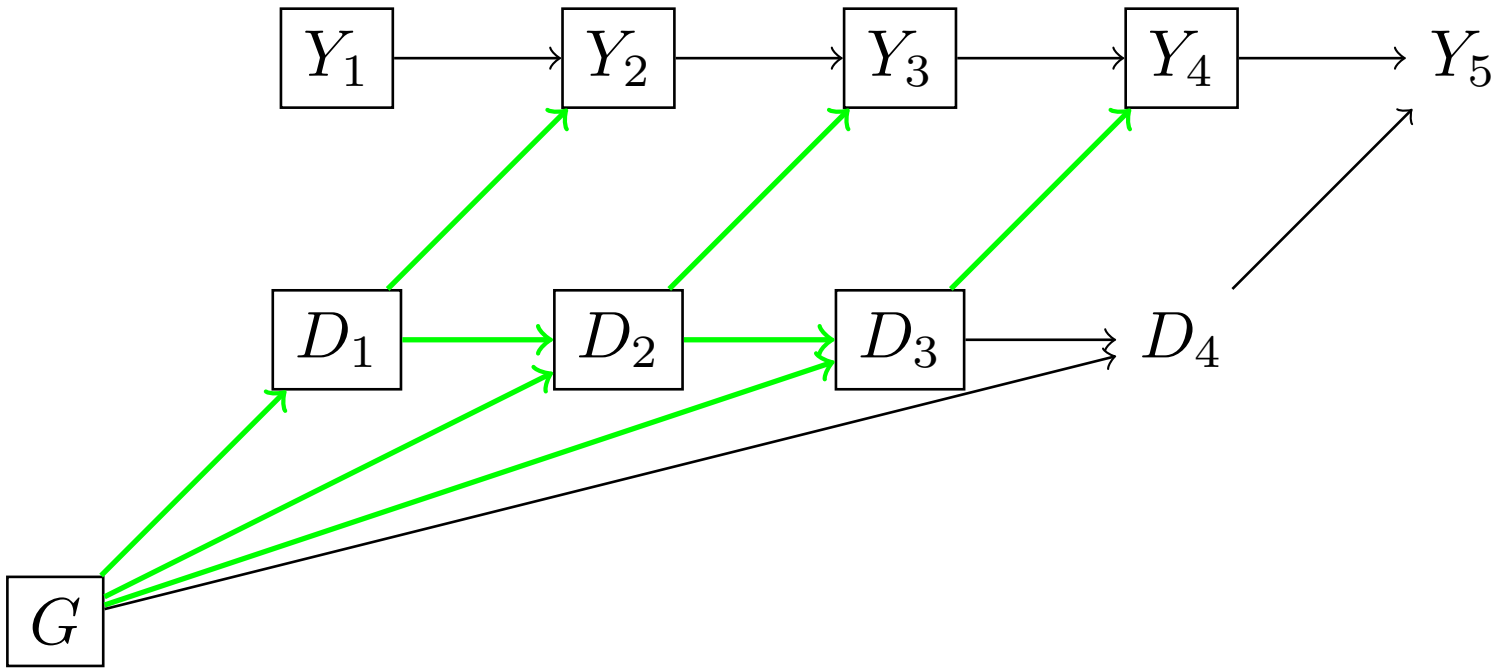
2. Mediation in time – visualization



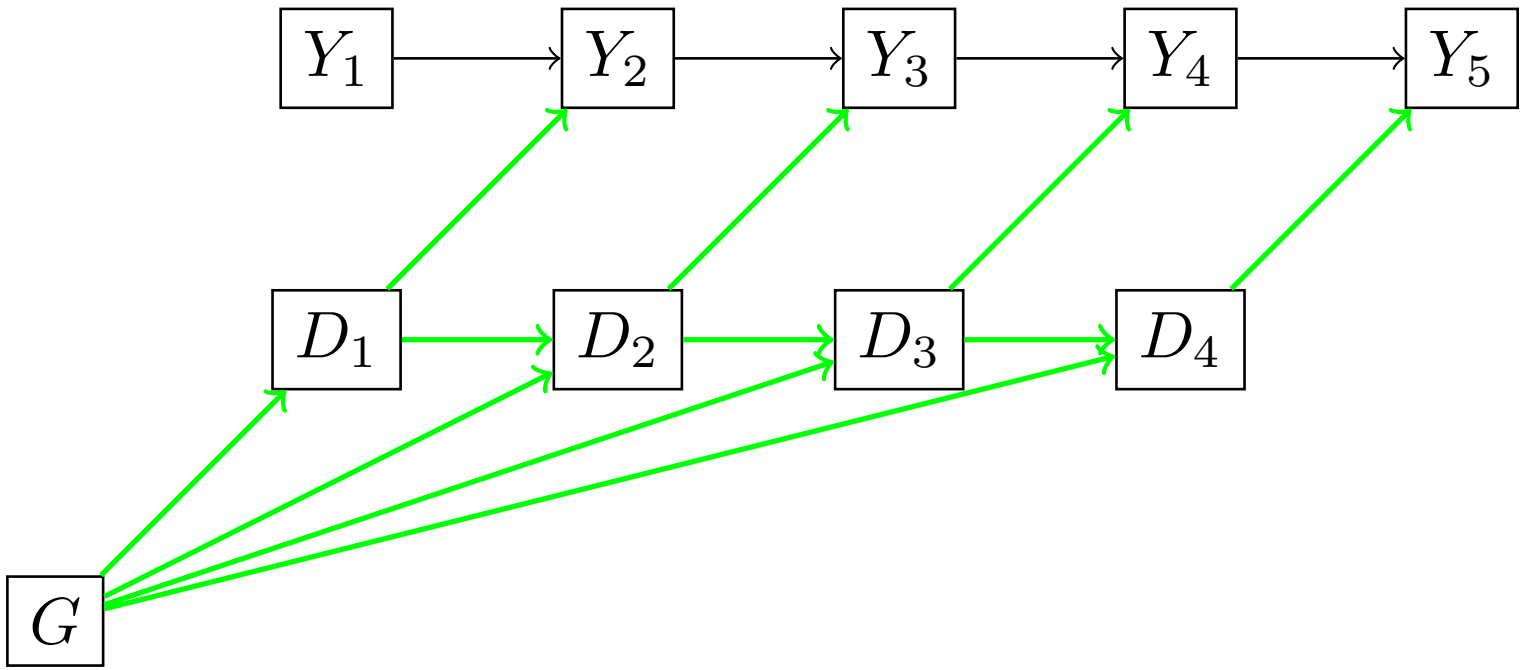
2. Mediation in time – visualization



2. Mediation in time – visualization



2. Mediation in time – visualization



2. Mediation analysis in time – use local independence?

"If a direct effect cannot reasonably be defined as a controlled or natural direct effect in the counterfactual sense because the required hypothetical manipulation of the mediator is inconceivable, then we can alternatively view these effects as being represented by flow in a dynamic system, so that the direct effect corresponds to the flow not passing through the mediator." [Aalen et al., 2012]

3. Fixed time points (landmark analysis)

Overcome the time problem by fixing time points

Fix t_0 and $l > 0$ and consider

$$\mathrm{P}\left(Y(t_0 + l) = 1 \mid D(t_0), T > t_0\right).$$

- Reasonable if t_0 and $t_0 + l$ denote some meaningful time of intervention and follow-up time, respectively.
- Lose information about what happens between t_0 and $t_0 + l$.

3. Fixed time points (landmark analysis)

Overcome the time problem by fixing time points

Fix t_0 and $l > 0$ and consider

$$\mathrm{P}\left(Y(t_0 + l) = 1 \mid D(t_0), T > t_0\right).$$

- Reasonable if t_0 and $t_0 + l$ denote some meaningful time of intervention and follow-up time, respectively.
- Lose information about what happens between t_0 and $t_0 + l$.

Local independence

- No need to fix time points.
- Better to capture "mechanistic workings"?

Relevance of local independence (summary)

1. Model of temporal dependence $\rightarrow$ Better way to compare (stopped) processes. Use obtained clusters as sub-diagnoses of HF.
2. An approach to mediation in a dynamical system $\rightarrow$ Answer questions about causal pathways for HF.
3. "Dynamic" point of view $\rightarrow$ Allows a more "exploratory" approach.

Major challenge

Most likely it will *not* hold that $G \not\to Y \mid D$. Then what?

Major challenge

Most likely it will *not* hold that $G \not\to Y \mid D$. Then what?

Effect estimation

Could we construct a good measure for the "strength" of the dependence? Should this measure be time-dependent? How should it be interpreted?

Major challenge

Most likely it will *not* hold that $G \not\to Y \mid D$. Then what?

Effect estimation

Could we construct a good measure for the "strength" of the dependence? Should this measure be time-dependent? How should it be interpreted?

Time periods of dependence and independence?

Would it be more informative to try and identify age spans during which dependence is present?

Other (major) challenges

Missing data

Missing observations, censoring, selection bias, discretization.

Interpretation

Does the concept of local independence lead to models with a clear interpretation? Does it bring us closer to causal interpretations – and if so, in what sense?

High-dimensional data

Genes...

References

O. O. Aalen. Dynamic modelling and causality. *Scandinavian Actuarial Journal*, 1987(3-4):177–190, 1987.

O. O. Aalen, Ø. Borgan, N. Keiding, and J. Thormann. Interaction between life history events. nonparametric analysis for prospective and retrospective data in the presence of censoring. *Scandinavian Journal of Statistics*, pages 161–171, 1980.

O. O. Aalen, K. Røysland, J. M. Gran, and B. Ledergerber. Causality, mediation and time: a dynamic viewpoint. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(4):831–861, 2012.

O. O. Aalen, K. Røysland, J. M. Gran, R. Kouyos, and T. Lange. Can we believe the dags? a comment on the relationship between causal dags and mechanisms. *Statistical methods in medical research*, 25(5):2294–2314, 2016.

V. Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):245–264, 2008.

A. B. Jensen, P. L. Moseley, T. I. Oprea, S. G. Ellesøe, R. Eriksson, H. Schmock, P. B. Jensen, L. J. Jensen, and S. Brunak. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nature communications*, 5(1):1–10, 2014.

S.-H. Lin, J. G. Young, R. Logan, and T. J. VanderWeele. Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Statistics in medicine*, 36(26):4153–4166, 2017.

S. W. Mogensen, N. R. Hansen, et al. Markov equivalence of marginalized local independence graphs. *The Annals of Statistics*, 48(1):539–559, 2020.

T. Schweder. Composable markov processes. *Journal of applied probability*, 7(2):400–410, 1970.

Questions / discussion

Local independence – mathematical definition

$X(t)$ is a multivariate stochastic càdlàg process

$$X(t) = (X^1(t), X^2(t), \ldots, X^k(t)), \quad t \in [0, T],$$

on $[0, T]$. For any $C \subset V := \{1, \ldots, k\}$ let $\mathcal{F}_t^C$ denote the completed and right continuous version of $\sigma\{X_s^c : s \leq t, c \in C\}$.
For $\beta \in V, C \subset V$, let $\Lambda^{C,\beta}$ denote the compensator of $\mathbb{E}[X^\beta(t) \mid \mathcal{F}_t^C]$, i.e., $\Lambda^{C,\beta}$ is a $\mathcal{F}_t^C$-predictable process and

$$\mathbb{E}[X^\beta(t) \mid \mathcal{F}_t^C] - \Lambda^{C,\beta}$$

is a martingale.
Then for $A, B, C \subset V$, X^B is said to be *locally independent* of X^A given X^C if there exists an $\mathcal{F}_t^C$-predictable version of $\Lambda^{C \cup A,\beta}$ for all $\beta \in B$. This is written as

$$A \not\to B \mid C.$$