Causal Reasoning for Events in Continuous Time: A DecisionTheoretic Approach

Vanessa Didelez

School of Mathematics

University of Bristol

with Kjetil Røysland & Odd Aalen (Oslo), Theis Lange (Copenhagen)

UAI — Advances in Causal Inference — July 2015

Based on:

Røysland, K., Didelez, V., Nygard, M., Lange, T., Aalen, O.O. (2015?). Causal reasoning in survival analysis: re-weighting and local independence graphs. Submitted.

Combining and extending:

Dawid, A.P., Didelez, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision theoretic overview. Statistics Surveys, 4:184-231.

Didelez, V. (2006). Asymmetric separation for local independence graphs. In Proc. of 22nd Conference in Uncertainty in Artificial Intelligence, AUAI Press.

Didelez, V. (2008). Graphical models for marked point processes based on local independence. JRSSB, 70(1):245-264.

Røysland, K. (2011). A martingale approach to continuous time marginal structural models. Bernoulli, 17(3):895-915.

Røysland, K. (2012). Counterfactual analyses with graphical models based on local independence. Annals of Statistics, 40(4):2162-2194.

Overview

- Local Independence for Processes
- Local Independence Graphs and δ–Separation
- Notion of Causal Validity ($=$ Extended Stability)
- Re–Weighting for Processes
- Identification Results
- Example: Norwegian cancer screening study

Local Independence

A Notion of Dynamic Independence among Processes

Notation

Multi-state process $Y(t)$ / marked point process (MPP);

$\Rightarrow$ represented by (collection of) counting processes $\{N_j(t)\}$ for each type of state change;

Note: will *not* always clearly distinguish $Y(t)$ and $\{N_j(t)\}$.

Mostly: time-to-event, e.g. $C =$ censoring time $\Rightarrow N^c(t) = I\{C \leq t\}$.

Aim: Graphical Representation of Dynamic Relations

For stochastic processes $X(t), Y(t), Z(t)$, represent and investigate (conditional) independencies of the type

$$\text{present of } X \perp\!\!\!\perp \text{ past of } Y \mid \text{ past of } (X, Z)$$

or (a little) more formally

$$X(t) \perp\!\!\!\perp \mathcal{F}_{t^-}^{Y} \mid \mathcal{F}_{t^-}^{X,Z}$$

where $\mathcal{F}_t^k$ filtrations, i.e. information becoming available over time.

Note: Asymmetric independence!

Links to Other / Earlier Work

Granger (1969): "Granger non–causality" for time series

Schweder (1970): "**Local independence**" for Markov processes

$\longrightarrow$ extended by Aalen (1987) and Didelez (2006, 2007, 2008)

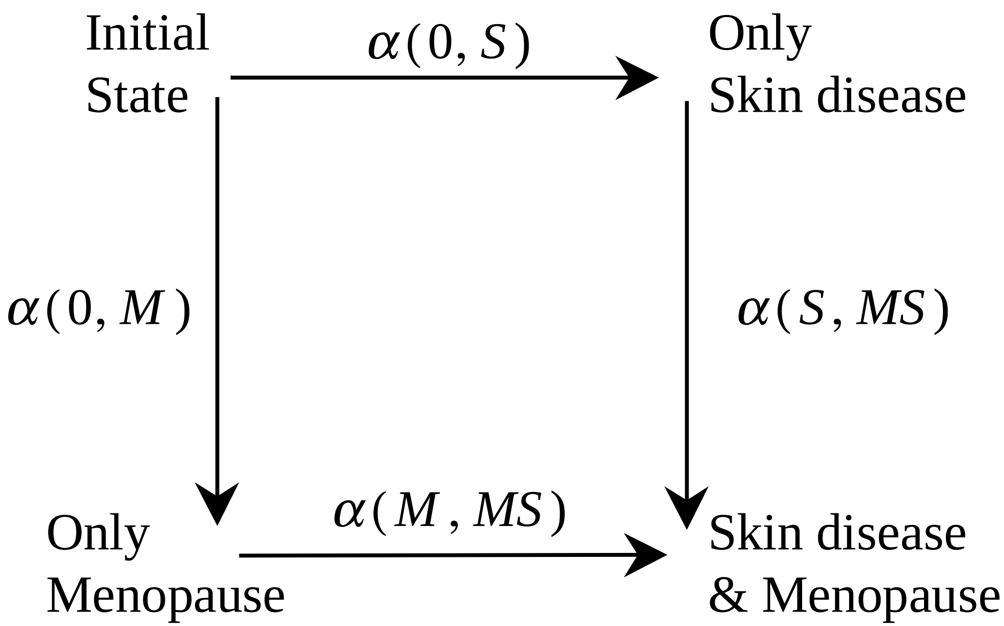
Graphical representations:

Eichler (2000, 2006) for time series.

Nodelman et al. (2002, 2003) for Markov processes.

Idea of Local Independence

Example (cf. Aalen et al., 1980) — Transition graph



Note: no transition $0 \to MS$ (**composable** Markov process)

Local independence: $\alpha(0, M) = \alpha(S, MS)$ while $\alpha(0, S) < \alpha(M, MS)$

Sneak preview: **Local Independence Graph**

... quite simple:

V = vertices = events

E = edges / arrows = local dependence

Bivariate case: dependence $\leftrightarrow$ independence

Multivariate case: conditional local in/dependence

Local Independence for Markov Processes

Consider for example Markov process with three components

$\mathbf{Y}(t) = (Y_1(t), Y_2(t), Y_3(t))$

— states $\mathbf{y} = (y_1, y_2, y_3) \in \mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3$

— s.t. any change of state of $\mathbf{Y}$ is always within one component y_i

$\Rightarrow$ e.g. transition (y_1, y_2, y_3) to (y_1, y_2', y_3) has intensity $\alpha_2(t; (\mathbf{y}, y_2'))$.

Local Independence: e.g. $Y_1 \not\rightarrow Y_2 | Y_3$ iff transition intensities satisfy

$$\alpha_2(t; (\mathbf{y}, y_2')) \text{ independent of } y_1$$

for all $\mathbf{y} \in \mathcal{S}$ and $y_2 \neq y_2'$ and for all t.

Why Local Independence?

For *small* $h > 0$

$$\alpha_2(t; (\mathbf{y}, y_2')) \cdot h \approx P(Y_2(t+h) = y_2' | \mathbf{Y}(t) = \mathbf{y})$$

so that $Y_1 \not\rightarrow Y_2 | Y_3$ implies for small h

$$Y_2(t+h) \perp\!\!\!\perp Y_1(t) | (Y_2(t), Y_3(t)).$$

Note: Not true in general for any $h > 0$.

Local Independence for Multi-State Processes

Preliminaries

Multi–state processes (or MPP more generally), K components
$\mathbf{Y}_V(t) = (Y_1(t), \ldots, Y_K(t))$, $V = \{1, \ldots, K\}$.

Under mild regularity conditions: Doob–Meyer decomposition

$$Y_k(t) = \underbrace{\Lambda_k(t)}_{\text{predictable}} + \underbrace{M_k(t)}_{\text{martingale}}$$

where $\Lambda_k(t)$ predictable based on history $\mathcal{F}^V{}_{t-}$ of whole $\mathbf{Y}^V$.

Note: will always assume intensity processes $\lambda_k(t)$ exist.

Local Independence for Multi-State Processes

...Doob–Meyer decomposition

$$Y_k(t) = \Lambda_k(t) + M_k(t)$$

where $\Lambda_k(t)$ predictable based on history $\mathcal{F}_{t-}^V$ of whole $\mathbf{Y}^V$.

Local independence: $Y_j \not\to Y_k | \mathbf{Y}_{V \setminus \{j,k\}} \quad \Leftrightarrow$

$\Lambda_k(t)$ (or $\lambda_k(t)$) is the same if information on past of Y_j is omitted

i.e. $\Lambda_k(t)$ is $\mathcal{F}_t^{-j}$–measurable.

Aside: Independent Censoring

Let $Y(t), X(t)$ be marked point processes, $C(t)$ indicates censoring.

Censoring is **independent** for Y given X if $C \not\rightarrow Y|X$.

Often: **independent** if $C \not\rightarrow (X, Y)$.

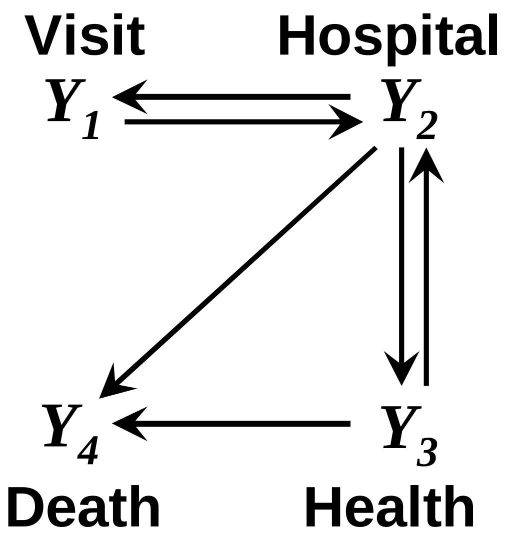
Example: violated if 'common cause' for censoring and event

Local Independence Graphs

and δ–Separation

Local Independence Graphs

$G = (V, E)$
V = nodes = components of process;
E = arrows = local dependence.



Graphs: directed, possibly cyclic, possibly two edges between pair of nodes.

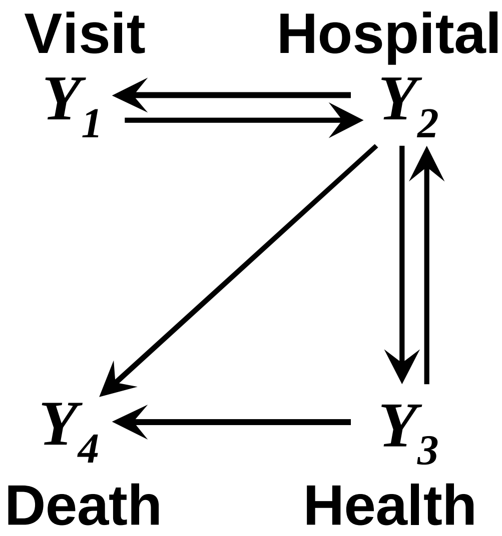
Note: $\text{pa}(k) \cap \text{ch}(k) \neq \emptyset$ possible.

Example: Home visits by nurses to elderly:

$Y_1(t)$ home visits by nurses at 'regular' intervals, increased rate only after hospitalisation, $Y_2(t)$ hospitalisation, $Y_3(t)$ health status, $Y_4(t)$ death.

Pairwise Dynamic Markov Property

Definition: $(j,k) \notin E \Rightarrow Y_j \not\rightarrow Y_k | \mathbf{Y}_{V\setminus\{j,k\}}$.



Hence, can see from graph: $Y_1 \not\rightarrow Y_4 | (Y_2, Y_3)$.

Want to know if $Y_1 \not\rightarrow Y_4 | Y_2$?

I.e. does Y_2 alone 'separate' Y_1 from Y_4?

$\Rightarrow$ will call this δ–separation.

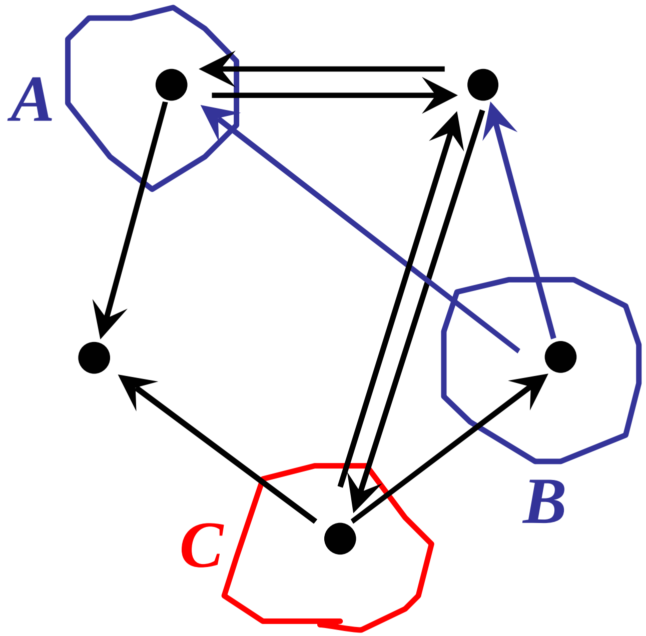
δ–Separation

Checking that C δ–separates A from B in a directed graph $G = (V, E)$:

Construct undirected graph in four steps

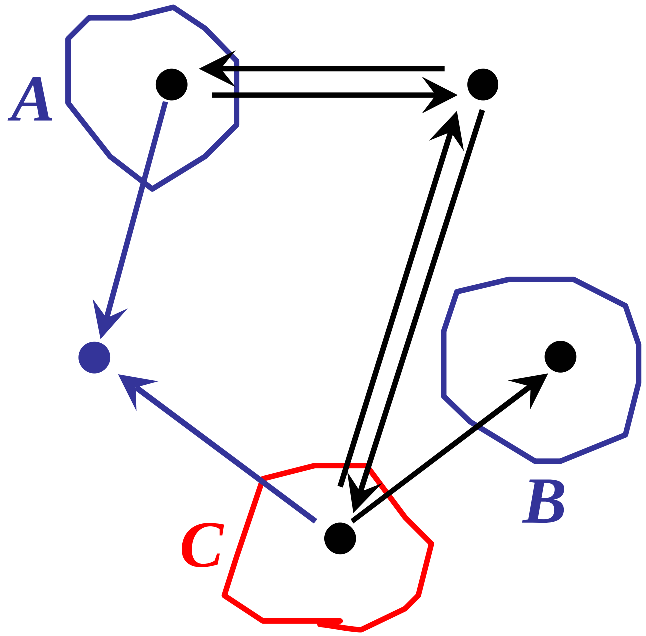
1. Delete edges starting in B;

(because: want to separate present of B from past of A, not interested in future of B)



δ–Separation

Checking that C δ–separates A from B in a directed graph $G = (V, E)$:



Construct undirected graph in four steps

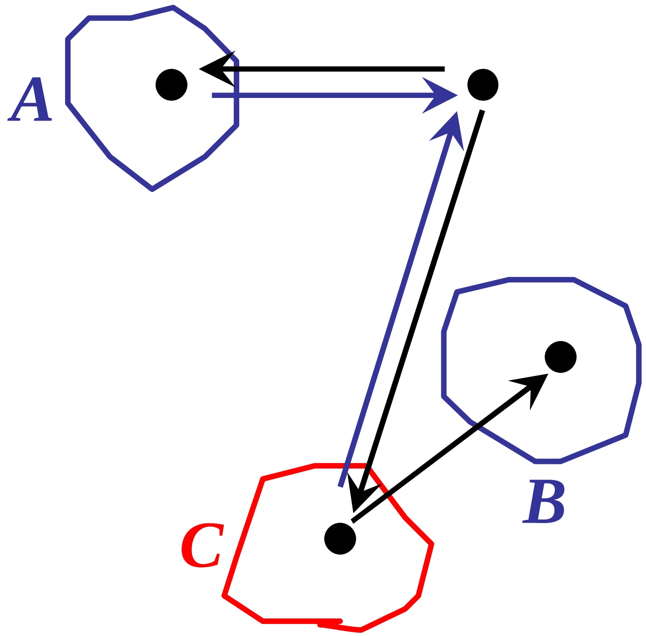
1. Delete edges starting in B;

2. Delete nodes not in An$(A \cup B \cup C)$;
   An(S) = set S and all its 'ancestors'

(1. and 2. are interchangeable.)

δ–Separation

Checking that C δ–separates A from B in a directed graph $G = (V, E)$:



Construct undirected graph in four steps

1. Delete edges starting in B;
2. Delete nodes not in $\mathrm{An}(A \cup B \cup C)$;
3. **'Marry' parents of common children;**

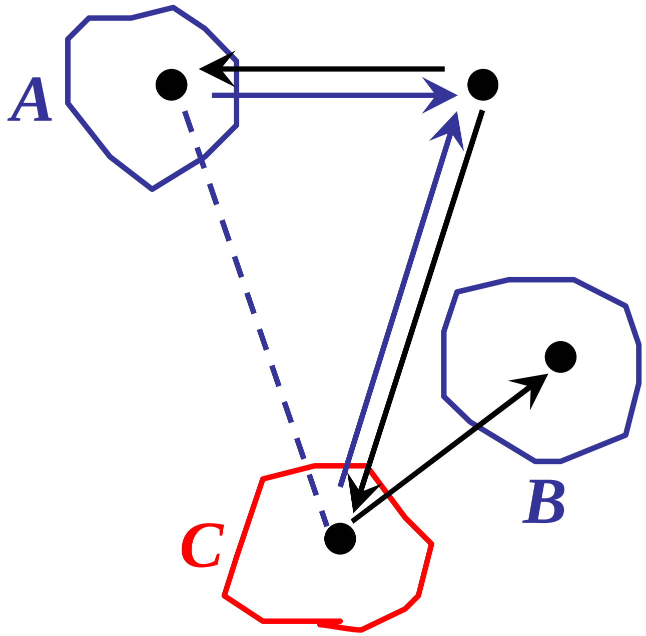
   (due to 'selection effect')

δ–Separation

Checking that C δ–separates A from B in a directed graph $G = (V, E)$:

Construct undirected graph in four steps

1. Delete edges starting in B;
2. Delete nodes not in $\mathrm{An}(A \cup B \cup C)$;
3. 'Marry' parents of common children;
4. **Make all edges undirected.**



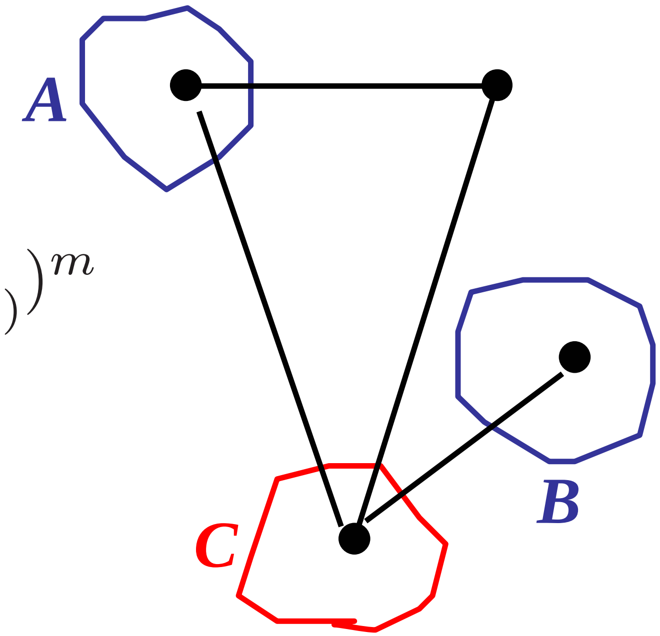
δ–Separation

Checking that C δ–separates A from B in a directed graph $G = (V, E)$:

Construct undirected graph in four steps.

In final undirected (moral) graph $(G^B_{\mathsf{An}(A \cup B \cup C)})^m$

check if C separates A and B in usual way.



Note: Still need to show that δ–separation meaningful in terms of local independence!

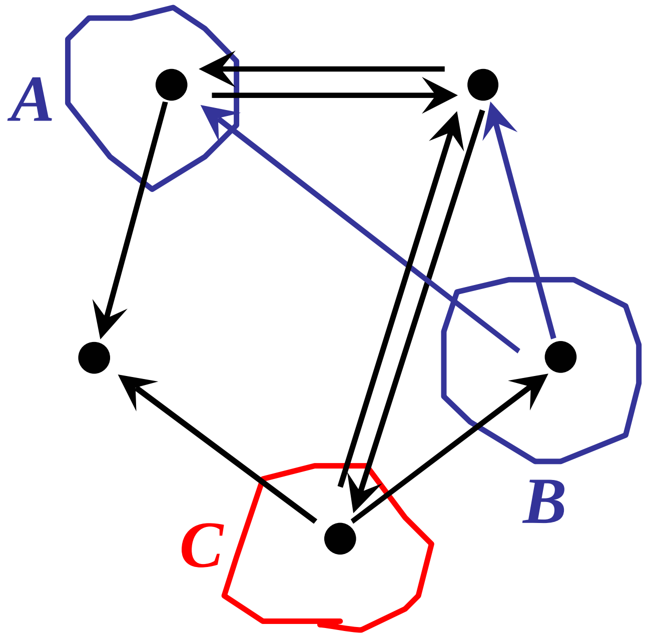
δ–Separation

Checking that C δ–separates A from B in a directed graph $G = (V, E)$:

Equivalently: every 'allowed' trail from A to B must be 'blocked' by C.

'allowed' $=$ must end in $\longrightarrow B$.

'blocked' $=$ same as in DAGs.



Global Dynamic Markov Property

Pairwise dynamic Markov property is (under regularity conditions) equivalent to

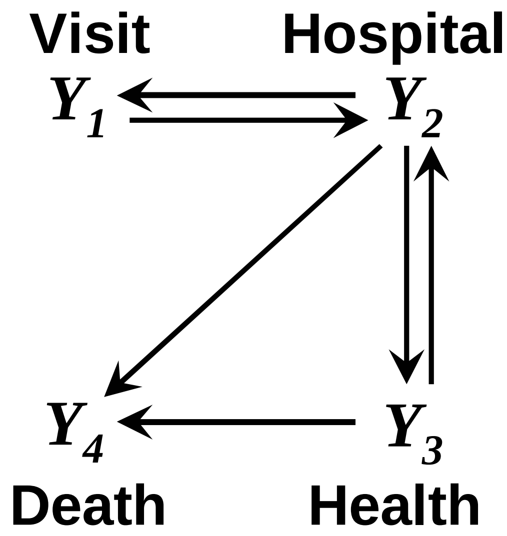
$$\text{whenever } C\ \delta\text{–separates } A \text{ from } B \text{ then } A \not\to B|C.$$

Proof: relies on asymmetric graphoid properties. (Didelez, 2006, 2008)

Home Visits Example

Back to initial example.

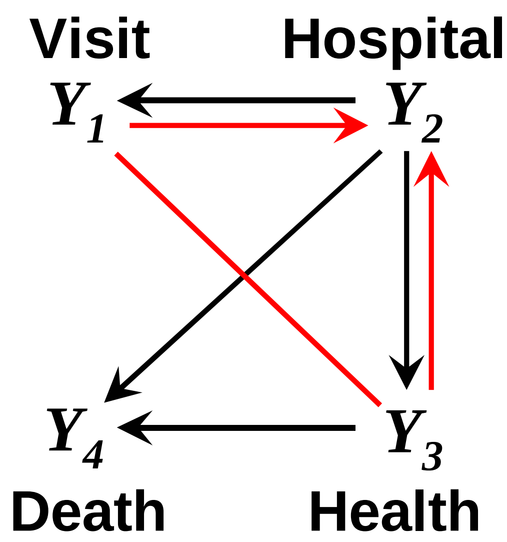
Does $Y_1 \not\rightarrow Y_4 | Y_2$?



Home Visits Example

Back to initial example.

Does $Y_1 \not\rightarrow Y_4 | Y_2$?



No!

A history of hospitalisation with prior home visit carries a different information on the health status than a history of hospitalisation without prior home visit.

Here, Y_3 time-varying confounder.

Comments

- Local Independence Graphs and δ–separation assume each process depends on its own past whatever other processes are included, hence, no 'self loops' shown.

- Meek (2014) extends this to allow a distinction between presence or absence of self-loops;
  – defines δ^*–separation
  – additional separations obtain in absence of self loop

Causal Interpretation

$\approx$ ‘Extended Stability’

Causal Validity

Analogous to many formulations of causality:

- Intervention: change intensity of one (or more) type(s) of event(s);
- Assume that in 'sufficiently' detailed system remaining intensities stay the same;
- Identification: what needs to be observed to estimate properties under intervened system?

$\Rightarrow$ Similar to 'Extended Stability' of Dawid & Didelez (2010).

Causal Validity

Joint density restricted to events before t: $\exists$ functionals $Z_1, \ldots, Z_K$ s.t. joint density is given by

$$\prod_{i=1}^{K} Z_i(\mu_i, t)$$

where μ_i local characteristics, e.g. $Z_i(\lambda_i, t) = \prod \lambda_i(s_r) \exp(-\int \lambda_i(s) ds)$.

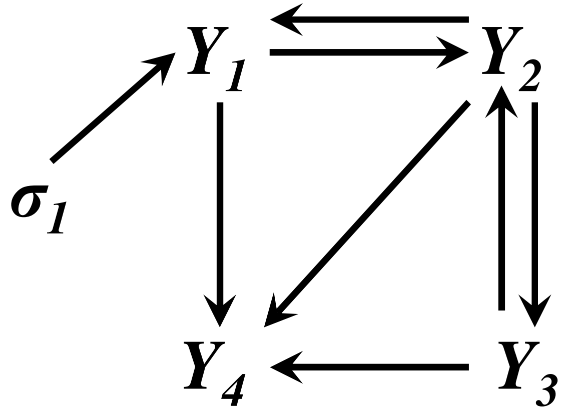
Note: each $\lambda_i(s)$ $\mathcal{F}^{\text{cl}(i)}$–measurable.

Causal Validity:

A specific intervention in i replaces μ_i by $\tilde{\mu}_i$, rest stays the same.

Note: can use intervention indicator σ_i in the spirit of Dawid (2002).

Intervention Indicator



Augmented graph G^σ with

indicator $\sigma_i \in \{o, e\}$, where

$\sigma_i = o$ means "observational regime"

$\sigma_i = e$ means "interventional regime",

i.e. change to $\tilde{\mu}_i$.

Absence of other edges with $\sigma_i \Rightarrow$ causal validity

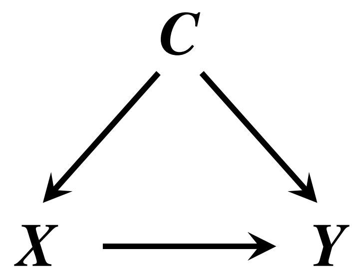
($=$ extended stability: Dawid & Didelez, 2010).

Example: as in home visits example, may want to know effect of changing rate of visits on survival; Y_3 unobserved.

Re–Weighting

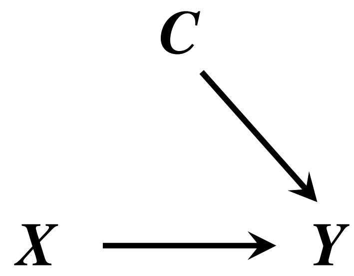
to Investigate Interventional Scenarios

Recall: Causal DAGs and IPTW



Observed data from:

$p(y, x, c) = p(y|x, c)p(x|c)p(c)$



Wanted: some aspect of *hypothetical*

$\tilde{p}(y, x, c) = p(y|x, c)\tilde{p}(x)p(c)$

Weights:

$$\tilde{p}(y, x, c) = p(y, x, c)W(x, c), \quad W(x, c) = \frac{\tilde{p}(x)}{p(x|c)}$$

and e.g.

$$\frac{1}{\#\{X_i = 1\}} \sum Y_i I\{X_i = 1\} W(X_i, C_i) \quad \text{estimates} \quad E(Y|\text{do}(X = 1))$$

Recall: Causal DAGs and IPTW

- Does hypothetical scenario make sense?
  $\Rightarrow$ subject matter; e.g. realistic interventions?

  C **identifies** $p(y \mid \text{do}(X = x))$.

- IPW $=$ change of measure

- ... related to importance sampling

- ... also used in financial maths for 'risk-neutral pricing';

- alternative: estimate $p(y|c, x)$, $p(c)$, substitute – "g-computation".

Re-Weighting for MPPs

Given: MPP with local independence graph G, causally valid wrt. Y_i.

P original model, $\tilde{P}$ model under intervention, $\tilde{P} \ll P \Leftrightarrow$

$$W(t) := \prod_{s \leq t} \left(\frac{\tilde{\lambda}_i(s)}{\lambda_i(s)} \right)^{\Delta N_i(s)} \exp\left(\int_0^t \lambda_i(s) - \tilde{\lambda}_i(s) ds \right)$$

uniformly integrable

e.g. $\lambda_i(s), \tilde{\lambda}_i(s)$ not too different, e.g. unif. bounded (Girsanov)

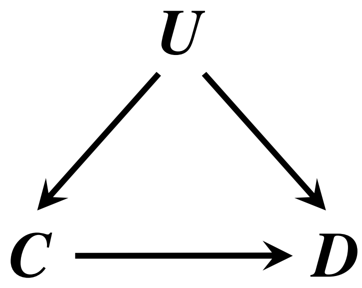
Note:

to be useful make $\tilde{\lambda}_i(t)$ measurable wrt. observed & relevant processes.

Causal Validity and Censoring

Folklore: "independent censoring yields K-M curves as if censoring had been prevented."

Counterexample: processes / times D, C, U; 'common cause' U ignored



K-M estimand in marginal model different from K-M estimand in model where 'intervention' prevents censoring.

Note: $C \to D$ often implausible if suitable (latent) processes included.

Identification

Eliminate Censoring and Modify ‘Treatment’ Process

Notation

Processes $V = V_0 \cup L \cup U \cup \{N^c, N^x\}$

$V_0(t)$ outcome processes of interest

$L(t)$ observed but not of interest

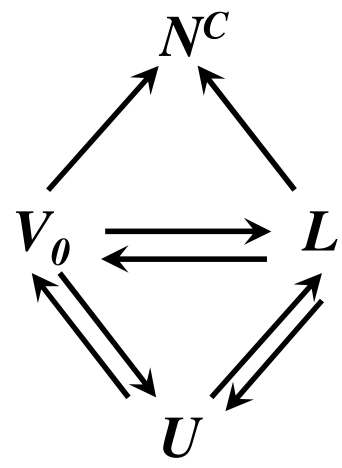
$U(t)$ unobserved

$N^c(t)$ counting process for censoring

$N^x(t)$ or $X(t)$ 'treatment' process

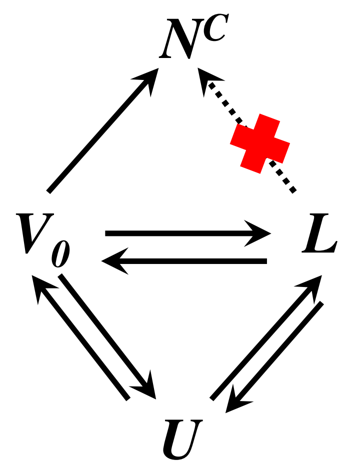
Causal Validity and Censoring

Processes $V = V_0 \cup L \cup U \cup \{N^c\}$



Consider P causal wrt N^c;
intervention replacing λ^c by
(V_0-predictable) $\tilde{\lambda}^c$ yielding $\tilde{P}$; $\tilde{P} \ll P$.

Theorem 1 **(Røysland et al, 2015):**
de$(N^c) = \emptyset$; if $U \not\to N^c|(L, V_0) \Rightarrow$ (censoring) weights $W(t)$ and V_0-intensity of any $N \in V_0$ identified without U.



Stabilised weights

$W(t) := \exp(\int_0^t \lambda^c(s) - \tilde{\lambda}^c(s)\, ds)$

$\Rightarrow$ 'deletes' arrows from L to N^c.

Causal Validity and 'Treatment'

Let processes $V = V_0 \cup L \cup U \cup \{N^x, N^c\}$

will ignore censoring N^c here

Consider P causal wrt N^x;

intervention replacing λ^x by (V_0-predictable) $\tilde{\lambda}^x$ yielding $\tilde{P}$; $\tilde{P} \ll P$.

Theorem 2 (Røysland et al, 2015):

If $U \not\to N^x | (L, V_0) \Rightarrow$ (treatment) weights $W(t)$ and V_0-intensity of any $N \in V_0$ identified without U.

More General: Simple Stability

Conjecture 3 **(Røysland & Didelez, 2015?):**
Using augmented local independence graph G^{σ}, a more general identifying criterion is

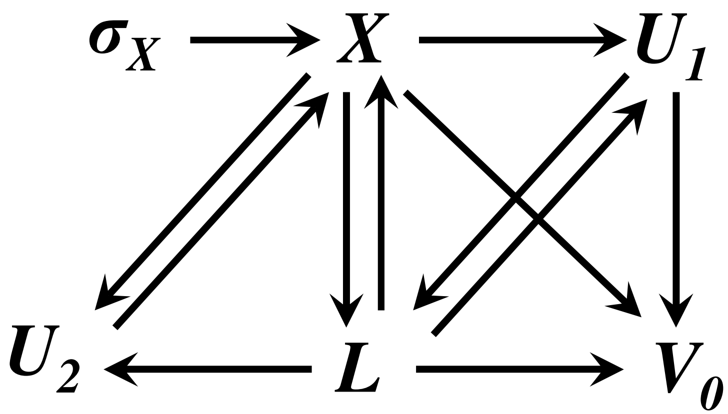
$$\sigma_X \not\rightarrow (V_0 \cup L) \mid X$$

$\approx$ **simple stability** (Dawid & Didelez, 2010)
$\Rightarrow$ is implied by Theorem 2

$\Rightarrow$ is also implied by 'sequential irrelevance': $U \not\rightarrow (V_0 \cup L) \mid X$.

Simple Stability — Example

$U = (U_1, U_2)$, check $\sigma_X \not\to (V_0 \cup L) \mid X$!



Crucial: no edges $U_1 \longrightarrow X$, $U_2 \longrightarrow (L, V_0)$, $U_1 - - - U_2$

Simple Stability ctd.

Pro: no need to refer to U

Con: does not refer to U...

Simple Stability ctd.

Pro: no need to refer to U

i.e. characterises a wide range of identifiable situations

Con: does not refer to U…

i.e. does not give intuition about what kind of U prevents identification

Application

Cancer Screening Process in Norway

Motivation: Cancer Screening *Process* in Norway

Norway: cervical cancer screening,
– women aged 25-69,
– every 3 years

7% "inconclusive" $\Rightarrow$ "triage": follow-up cytology and HPV-tests

$\Rightarrow$ (1) referred to invasive diagnostic procedure

(2) more tests "soon"

(3) return to 3-year testing

Motivation: Cancer Screening *Process* in Norway

HPV-tests: 3 types (brands): A/B, and C.

Type C:
negative test results more often followed by cytology = "lesions / worse"
$\Rightarrow$ unsuitable HPV-test?

But,

Type C: also subject to more frequent / sooner testing due to manufacturer's recommendation (twice as many in the first year)

Note: Government withdrew funding for company C's tests; company C was going to sue...

Target of Inference

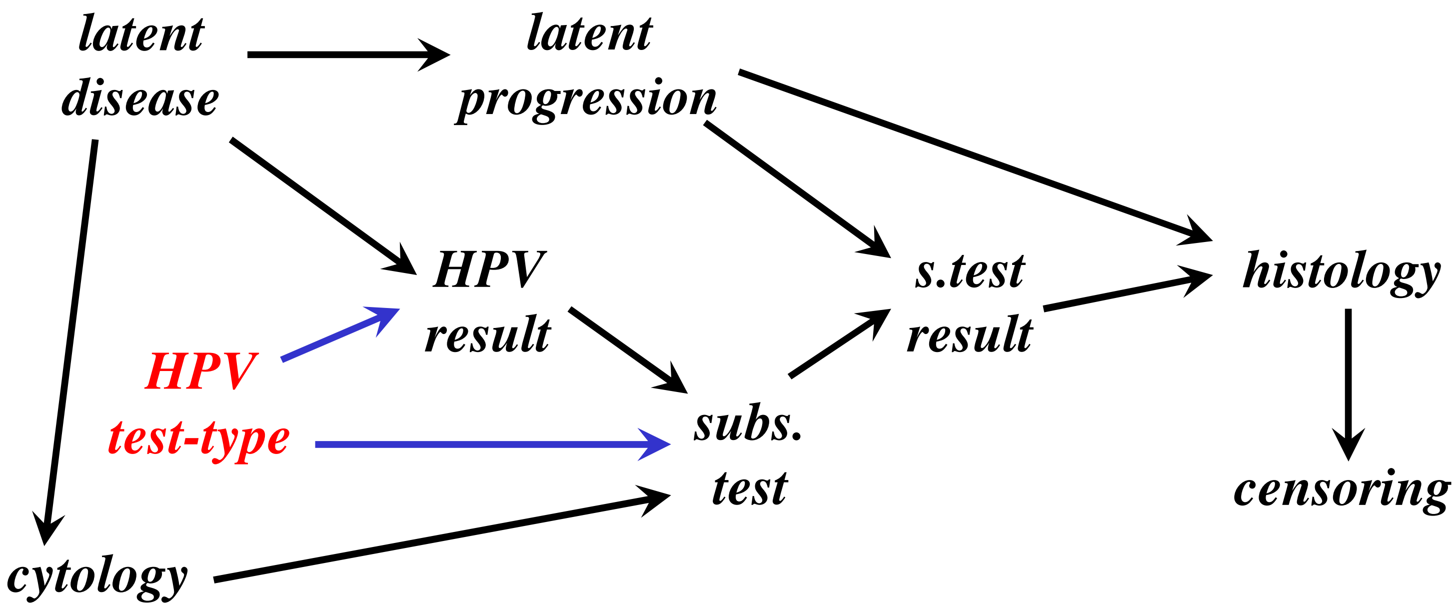
Comparison of incidents of alarming cytology results (CIN2+) after negative HPV-test from A/B versus C in *hypothetical scenario* where test-type C has subsequent testing as (in)frequent as types A/B.

$\Rightarrow$ replace hazard rate α^C by $\alpha^{A/B}$ for ‘subsequent testing’;

$\Rightarrow$ then look at re-weighted system;

$\Rightarrow$ all conditional on ‘in triage’ and ‘negative first HPV’ result.

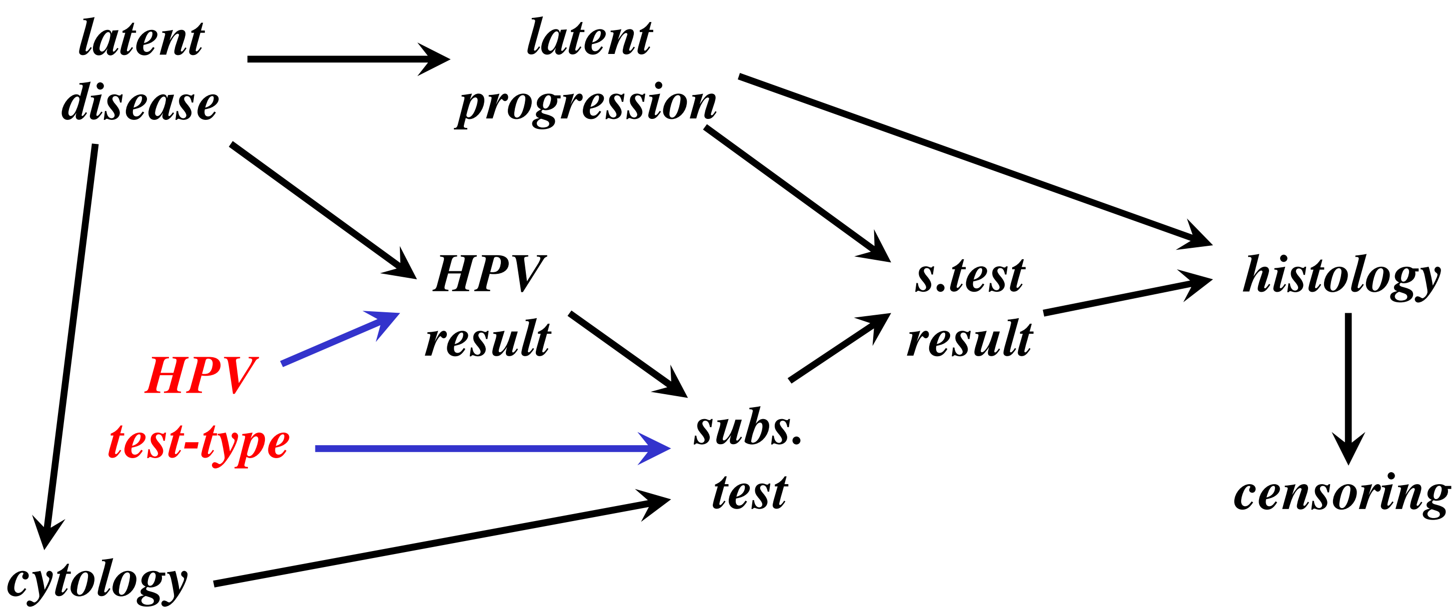
Assumptions



Censoring: registry data; censored at first CIN2+ result / end follow-up

$\Rightarrow$ Theorem 1 satisfied.

Assumptions

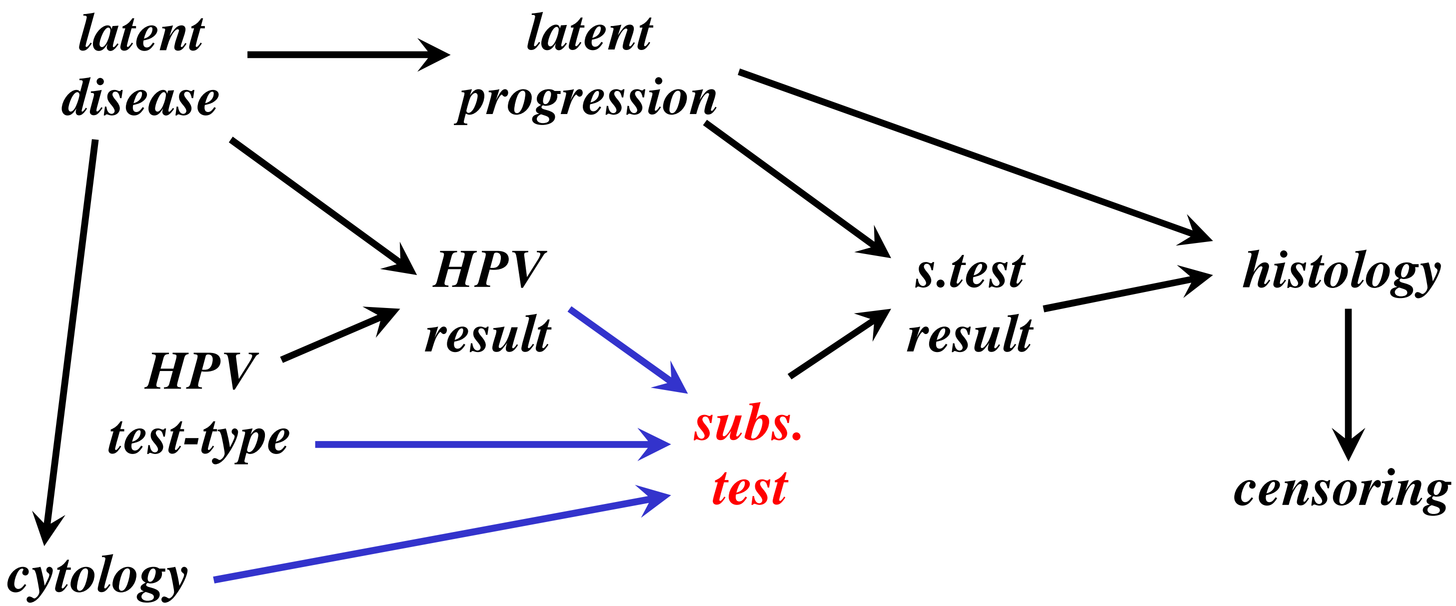


HPV-test type (A/B versus C) like ‘randomised’

Tests work differently ⇒ potentially different results

Type C known to be followed by more subsequent testing

Assumptions

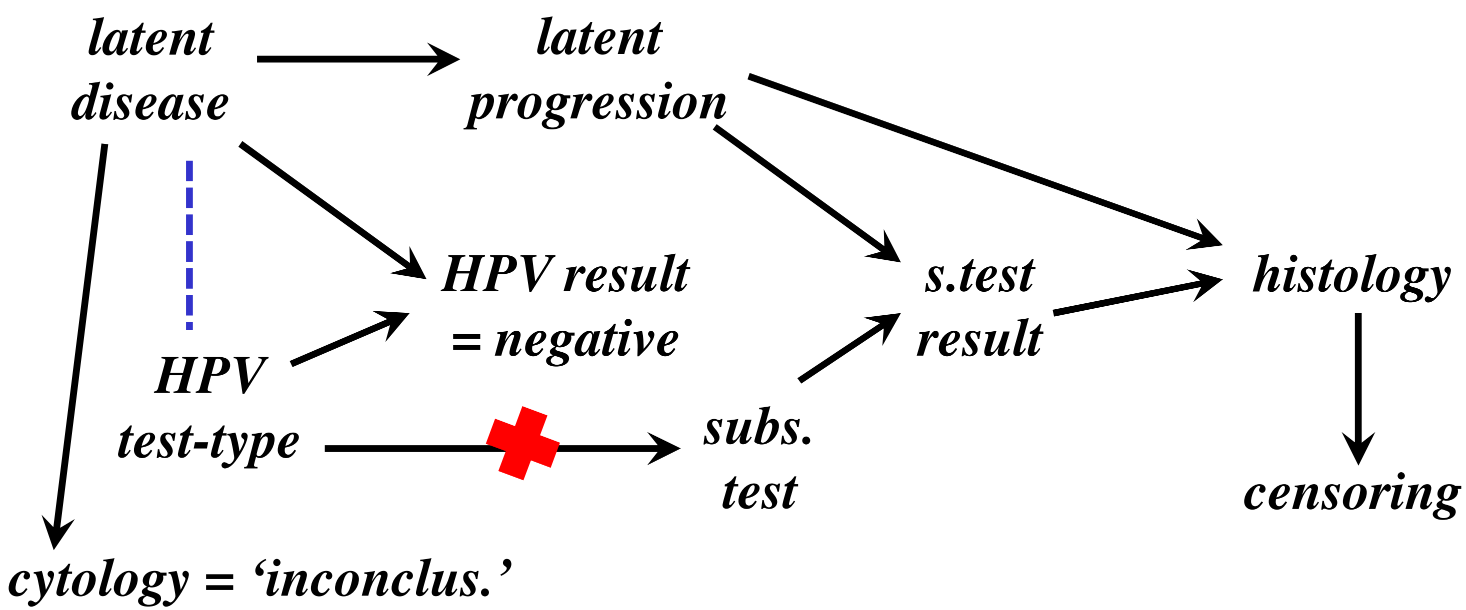


Subsequent testing follows protocol.

This is the process whose intensity we want to replace $= N^x$.

Note:
HPV test-type *not* regarded as treatment; but indirectly predictive.

Assumptions



Want to make subsequent testing independent of test-type.

Theorem 2 satisfied: subsequent testing loc. independent of latent variables / processes given observed processes.

Weights

T_j = 1st time individual j subject to subsequent testing.

Likelihood ratio in type-C group:

$$W^j(t) = \left(\frac{\alpha^{A/B}(T_j)}{\alpha^C(T_j)}\right)^{I\{T_j \geq t\}} \exp \int_0^{t \wedge T_j} \alpha^C(s) - \alpha^{A/B}(s)\, ds$$

Nelson-Aalen estimates $\hat{A}^C(s)$ and $\hat{A}^{A/B}(s)$ of cum. hazards.

$\Rightarrow$ smooth with splines, differentiate, plug-in to estimate $W^j(t)$.

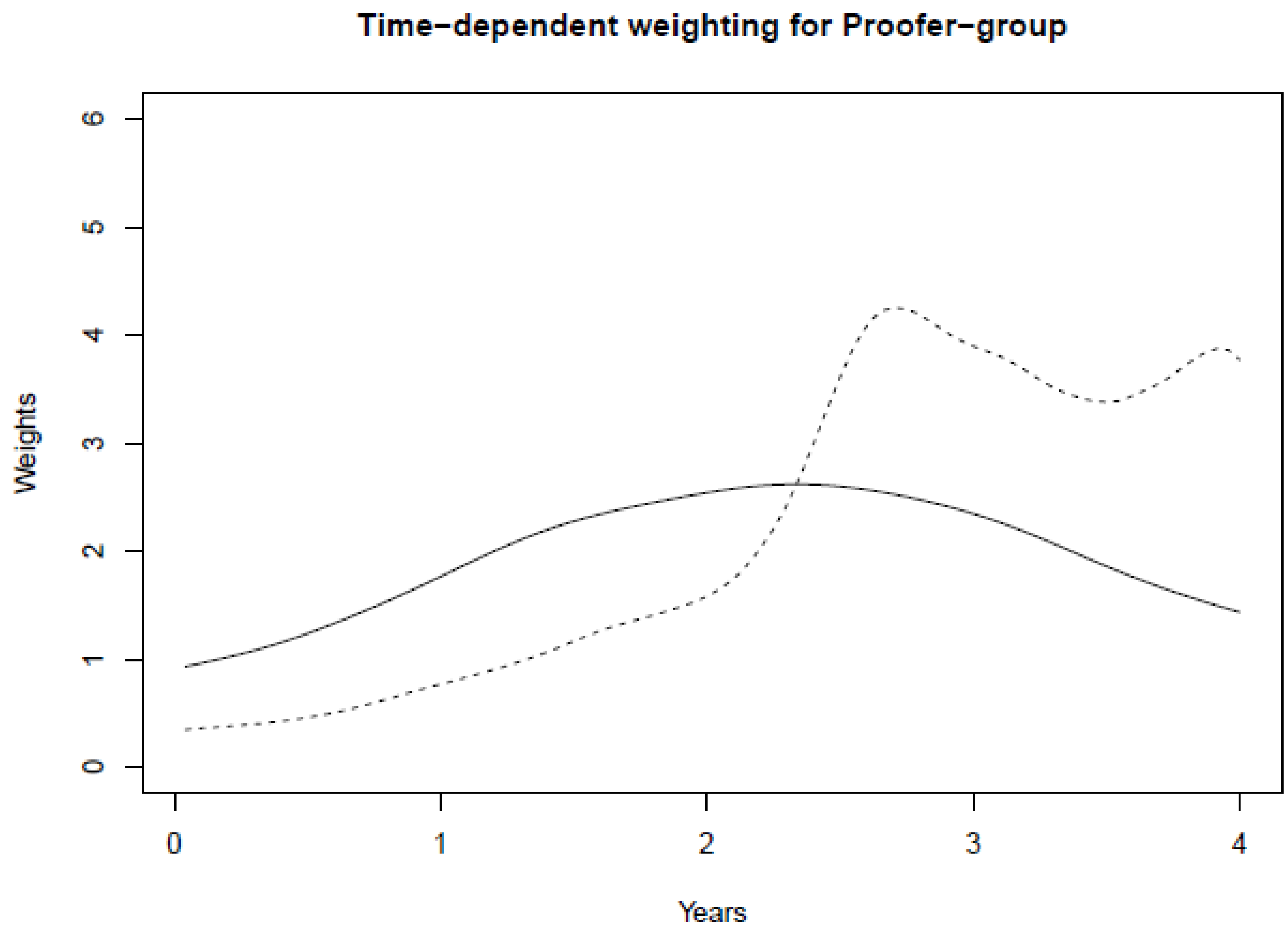
Re-weighted K-M; events and 'at' risk

$$\hat{N}(t) = \sum_j \int_0^t \hat{W}^j(s-)\, dN^j(s), \qquad \hat{Y}(t) = \sum_j \hat{W}^j(t-) Y^j(t).$$

Weights

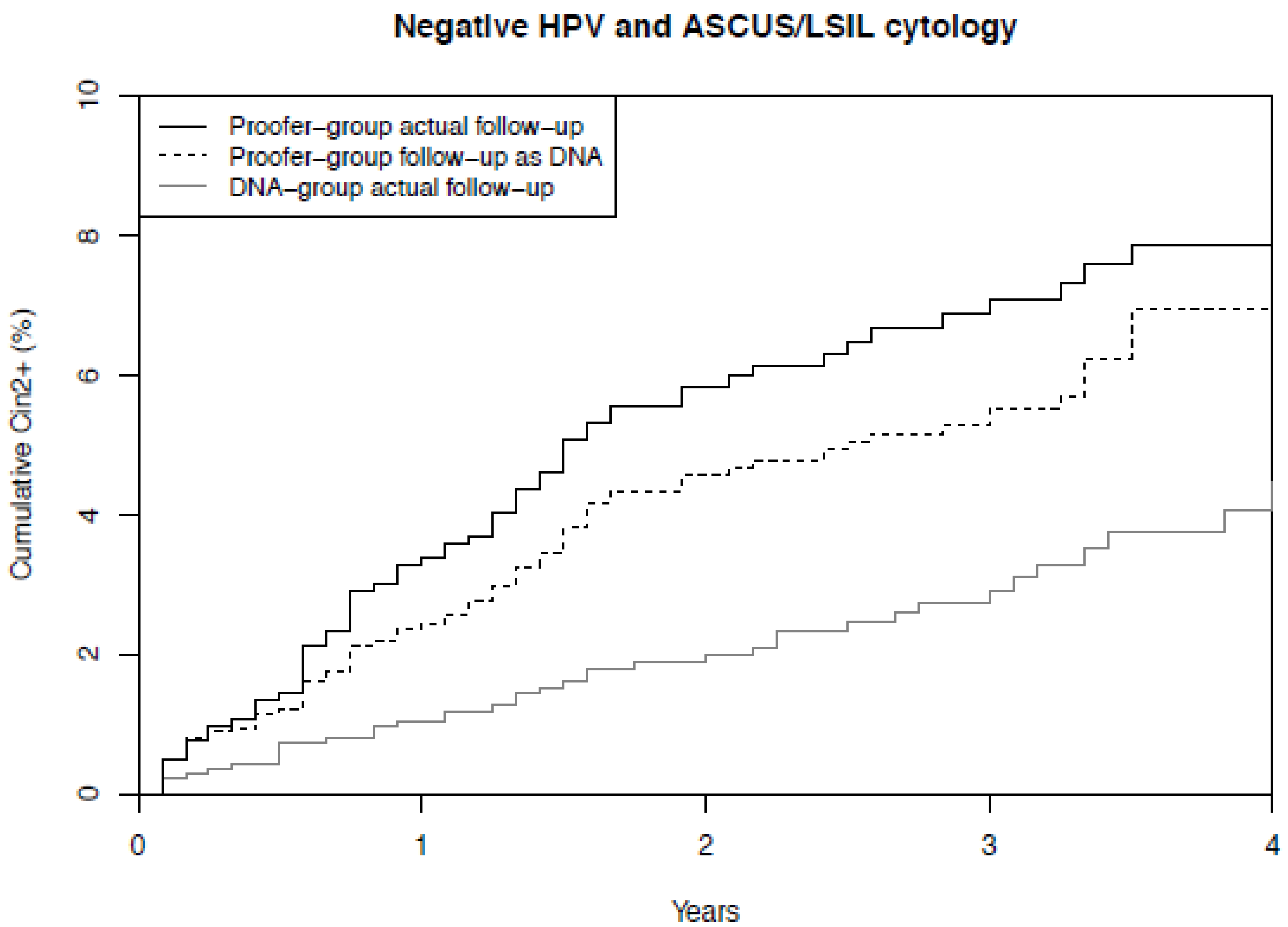
solid line: before test

dotted line: after test



Results

Re-weighted log-rank test: p-value=0.004.



Summary

- Local independence graphs useful to represent dynamic dependence structure in event histories.
- δ–separation suitable to investigate which independencies are preserved under marginalisation.
- Extension to causal reasoning: parallel to decision theoretic approach advocated by Dawid (2002); use intervention indicator for more general identification result $\Rightarrow$ t.b.continued
- Need to think 'causally' about censoring.
- Application to Norwegian cancer screening programme.
  Company C has withdrawn their law suit against government...
- For more complex applications / general processes, practical details to be worked out.