Causal inference and survival analysis – A new look at old tools

Symposium celebrating Niels Keiding
May 14-15, 2014
Odd O. Aalen
Department of Biostatistics, IMB
University of Oslo
and Cancer Registry of Norway
Congratulations to Niels!

- Niels has a very important scientific role. Modern survival analysis would not have been the same if it hadn’t been for Niels’ ability to pull together various people, various ideas, to see the possibilities and **the grand scheme of things**.

- Niels is **generous**, supportive of ideas and people. He is interested in other peoples views as well, not only his own.

- Niels is **enthusiastic** about science and ideas
My generation grew up with the notion that causality was a forbidden word. At least, not to be uttered by humble statisticians.

Then, entering medical statistics one realized that medical people expected contributions from you to understand causality. That’s why they paid your salary.

But, we had to wait for some decades before this was taken seriously and the notion of causal inference was coming up.
Pearl’s book: The great revelation

- **Doing versus seeing**: Pearl’s do-calculus has become a major tool in epidemiology and other fields. Extensive mathematical theory for calculating causal effects.

- **Causality is about interventions**: What happens when we intervene?
Intervention is fundamental in biostatistics

- The final aim of medical research is to intervene (either to treat or to prevent disease).
- When listening to talks and reading papers you should look for the interventions lurking behind.
- Important question: Most data are just “seeing”. Can we deduce “doing” from “seeing”?
  - The Nordic countries are known for their excellent registries. But these vast, and expensive, amounts of data are all seeing, no doing. So the issue of causal inference is fundamental.
The other view of causality: the mechanism

- **Interventionist**: What is the effect of an intervention? The fundamental question of causal inference
- **Mechanistic**: Understanding the inner workings of a system. Looking for an explanation.
  - The essence of natural science
  - Time is essential: cause-effect relationship developing over time
  - Understanding the mechanism can help you to choose the right intervention

© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com
Mechanisms and causality

- Assume there is some causal effect from an intervention.
- Then **this must have a relationship to the mechanistic structure of the system**. Something must produce the effect of the intervention. Presumably this is some mechanism, some process, operating in continuous time.
- The underlying causal structure is fundamental and exists in the absence of interventions (or humans).
  - “... signal sensing is more fundamental to the notion of causation than manipulation; the latter being but a crude way of stimulating the former in experimental setup. The mantra “No causation without manipulation” must be rejected.” Pearl, 2011
Entering the processes

- Mechanisms may be modeled as stochastic processes
- Counting processes, martingales etc. have a fundamental role in causal inference
- Dynamic covariates may be relevant (say number of previous jumps)
- One must be careful, the process may not be well-defined

\[ I : \lambda(t; x_t, \ldots) = \alpha(t) \exp(\beta x_t + \ldots) \]
\[ II : \lambda(t; x_t, \ldots) = \alpha(t) + \beta_t x_t + \ldots \]

where \( x_t \) is the number of previous jumps

Model I may explode at any time, while model II is well defined due to Lipschitz condition. Gjessing, Røysland et al 2010 give detailed results on this
Counting processes and causality

- How is the relation between processes and intervention?
- Intensity processes are projections of the present, or near future, on the past
- They may appear to “explain” what is happening, but there is no guarantee that they do so correctly
- Using the $do$-operator of Pearl, we may expect:

$$\lambda(t; x,...) \neq \lambda(t; do(x),...)$$

- For causality we need structural models, that stay correct under intervention (Pearl, Robins, Hernán)
Seeing vs. doing: A frailty effect

Comparing two treatment groups by hazard ratio. You might think that when the hazard ratio equals 1, then there is no difference in treatment effect. Hence, everybody could use the default treatment from this time. However, this intervention may yield a sharp change in the event rate in the group that changes treatment. The reason for this is a frailty component.

If we have a mechanistic (frailty) understanding, this enables us to predict the effect of an intervention.
Frailty must be understood as a causal concept: Collider effects and survival

- Hazard rates are defined by conditioning on survival. In the presence of multiplicative frailty this introduces collider effects.
- Backdoor path in red: Survival curves are “more causal” than hazard rates
Additive models and survival collider effects

- Assume structural equations that model the covariates in an additive model
  \[ X = \varepsilon_1, \quad Y = kX + \varepsilon_2, \quad Z = mX + rY + \varepsilon_3 \]
  with independent \( \varepsilon \)'s.
- Then the relationship between covariates survives the collider effect and stays unchanged.
- This a foundation of dynamic path analysis
  - Strohmaier, Røysland, Aalen
  - Related results of Martinussen and Vansteelandt
Two worlds: How can Pearls theory of causal DAGs be extended to processes?

- What is a causal directed acyclic graph (DAG)?
  - Network: Acyclic
  - Probabilistic part: It is a Bayesian network, implying network defined by conditional independence (Markov cond.)
  - Intervention part: It has a prescription for what happens when you intervene
    - Intervening on a node removes all incoming arrows
    - Bayesian network structure otherwise unchanged

- What is a causal local independence graph?
  - Network: Can have cycles (feedback)
  - Probabilistic part: Network defined by local independence
  - Intervention part: It has a prescription for what happens when you intervene
    - Intervening on local characteristic: All other local characteristics unchanged
Doob-Meyer decomposition

\[ dX(t) = U(t)dt + dM_X(t) \]

- The dependence on the past is carried by the local characteristic \( U(t) \) while the martingale (innovations) represents new random impulses from outside (noise).
- Examples:
  - \( U(t) \) is an intensity process of a counting process
Local independence – two processes

\[ dX(t) = U(t)dt + dM_X(t) \]
\[ dY(t) = V(t)dt + dM_Y(t) \]

- \( Y \) is **locally independent** of \( X \) at time \( t \) if
  1. \( V(t) \) is not dependent on the past of \( X(t) \)
  2. \( M_X(t) \) and \( M_Y(t) \) are orthogonal martingales

If 1. is not fulfilled we have local dependence.

Example: Local independence graph for HIV treatment

Local independence graph showing relationship between processes. Does not have to be acyclic, can include feedback.

The highly active anti-retroviral treatment (HAART) is a very efficient treatment which primarily works through suppressing the replication HIV virus.

The clinical status of the patient also determines start of treatment.

Numerous confounders can be corrected for.
Local independence graphs can be used in much the same way as Bayesian networks (DAGs) in the sense of Pearl

- $\delta$-separation (Didelez, 2007) corresponds to d-separation
- Backdoor paths may exist
- Collider effects can be studied
- Arrows can be removed by inverse probability weighting
- The model is causal if intervening on a local characteristic does not change other local characteristics in the model (Røysland et al, 2014)
  - This implies the no unmeasured confounder assumption
Causal inference and stochastic analysis


- This connects *causality* to martingale problems, stochastic differential equations and counting processes.
What is the meaning of the Kaplan-Meier estimator?

- We think we know it, but do we? What does it mean?
- One interpretation: It estimates the survival curve in a counterfactual world where there is no censoring
- But this is a causal statement
- Is independent censoring (in the dynamic survival analysis sense) enough to guarantee this?
- No, because independent censoring is not a causal assumption
Local independence graph modelling independent censoring

Independent censoring is a fundamental concept in survival analysis. It is naturally formulated as local independence, see Figure. (Røysland, Arjas). Time-dependent confounding has a similar interpretation. But this may not be causal.

treatment process: \( X \)
outcome process: \( Y \)
observed confounder process: \( L \)
censoring process: \( C \)
independent censoring
Røysland, counter example

- C-censoring, N-counting process, U-unmeasured confounder

- Upper graph: true relation

- Lower graph: relation between observed processes

- Details can be had from Røysland
Dependent censoring in (b) can be handled by inverse probability weighting, IPCW (Robins). Local independence graph gives a compact description.

If the model is causal, then inverse probability weighting gives a causally correct result.

Røysland, Didelez
Analyzing the effect of censoring

- Local independence graph to judge the effect of censoring
- The effects of unmeasured confounders can also be evaluated
- Røysland, Didelez, Lange, Aalen

C is $\delta$-separated from Y by X and L, hence censoring is independent when everything is observed. X alone does not -separate C from Y, and so the marginal analysis given by Kaplan-Meier plots does not give a causally valid comparison. But inverse probability weighting resolves this issue.

C blocks part of the causal effect. However, we can estimate the direct causal effect as well as the indirect effect passing through X.
The advantage of the local independence graph is **compactness**.
Can randomized trials be simulated from routinely collected clinical data?

- **Important** because increasingly large amounts of clinical data will be available
- **Difficult causal issue** because medical treatment is intimately connected with the underlying disease process. This is different from most risk factor epidemiology
- Fundamental ideas by Robins and co-authors
- Large HIV cohorts in the US, UK and Switzerland have been used as a testing ground for new methodology. We cooperate closely with the Swiss HIV cohort
- Results from several groups show that conclusions on treatment effects are feasible
Can we sort out the feedback and estimate the treatment effect? (Also called time-dependent confounding)

- Naïve estimation goes wrong
- Robin’s marginal structural model (MSM) is one solution
- This can be put in a counting process framework which yields a simple and general mathematical derivation

Assumptions:
- Positivity
- No unmeasured confounder
Time-dependent confounding

Local independence graph:
treatment process: $X$
outcome process: $Y$
observed confounder process: $L$
censoring process: $C$

Assume the model is causal – no unmeasured confounding

Røysland (Bernoulli 2011) uses local independence and counting process theory (Jacod, 1975) to prove how the marginal structural model (MSM) performs a reduction of the graph. Appropriate inverse probability weighting makes arrows disappear.
Swiss HIV cohort data

- An ongoing multi-center research project following up HIV infected adults aged 16 or older. 2161 individuals, observed over minimum 1 and maximum 92 months.

- Data goes from 1996 to September 2003. The data are organized in monthly intervals, with measures of CD4 count (measure of immune system status), viral load (HIV-1 RNA) and numerous other clinical information.

- This dataset has already been analyzed using MSMs [Sterne et al. 2005]

- We present MSM results together with those from an alternative procedure, named sequential Cox regression (Gran, Røysland, Didelez et al)
The sequential Cox approach

- *Mimic a sequence of randomized clinical trials (RCTs)* based on each time period (month) of treatment start
- *Average over all* mimicked RCTs to find an overall effect estimated by composite likelihood
- Consider only individuals starting treatment in a certain time interval as the *treatment group*. Analysis start at the starting point of this interval
- Individuals still not on treatment in this interval serve as the *control group* – if they start treatment on a later stage they get censored (*artificial censoring*)
- Dependent censoring? Solution: inverse probability of censoring weighting
- Adjust for covariates at baseline and at the start of the mimicked RCT (using for instance a Cox model)
Results from overall analysis

<table>
<thead>
<tr>
<th>Model Description</th>
<th>HR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unweighted Cox model, baseline and time dependent covariates</td>
<td>0.647</td>
<td>0.430-0.973</td>
</tr>
<tr>
<td>Unweighted Cox model, baseline covariates only</td>
<td>0.334</td>
<td>0.232-0.483</td>
</tr>
<tr>
<td>MSM</td>
<td>0.140</td>
<td>0.066-0.299</td>
</tr>
<tr>
<td>Sequential Cox</td>
<td>0.176</td>
<td>0.105-0.296</td>
</tr>
<tr>
<td>Censor weighted sequential Cox</td>
<td>0.165</td>
<td>0.079-0.343</td>
</tr>
</tbody>
</table>

Table 1: Estimated hazard ratio of HAART versus no treatment (confidence intervals for the sequential models estimated by Jackknife). The three first rows corresponds to results from Sterne et al.
Mediation

- The mediator concept is important, especially in biology and psychology. E.g. 2300 papers in the journal Nature using the word mediator.
- Mediation is often thought of as a mechanistic concept, describing some process happening over time.
- “… mediating variables are behavioral, biological, psychological, or social constructs that transmit the effect of one variable to another variable.” McKinnon et al, 2007

- Why are mediators of interest? Gives understanding of mechanism. Can be targeted for treatment interventions.
- In causal inference natural and controlled direct effects are defined counterfactually. Focusing on just a few variables, usually no process is considered.
What happens under discrete observation of a continuous process?

- One rarely observes the continuous process, usually one just has some observation at discrete time points. What is the effect of this?
- The network properties are not preserved. Regarded as a Bayesian network the time-discrete structure will fill up with arrows.
- There is a fundamental inconsistency between the continuous and discrete data.
- It is a problem that almost all theory for causal mediation takes place at the discrete level.
- Nature operates in continuous time (apart from quantum discretization).
Artificial direct effects in time-discrete graphs

- Assume the absence of direct effects at the process level. Then time-discrete observation will not give this picture, but will show apparent direct effects as well.

- The system will fill up with arrows under discretization.
  - Lange, Røysland, Gran, Aalen, Kouyos
Figure 6. Analysis of increments of CD4 and HEM based on repeated regression analyses with lagged and baseline values of RNA, CD4, and HEM as independent variables. Analyses shown for monthly, yearly, and tri-yearly values from left to right.
Emerging picture depends on frequency of observation

- Swiss HIV cohort data
- Considering the variables viral load (RNA), CD4 count, hemoglobin (HEM)
- Monthly intervals result in the believable mechanistic model:
  - Viral load → CD4 → Hemoglobin
- Coarse, time-discrete observations give a different view than finer observations
Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms
OO Aalen, K Røysland, JM Gran, R Kouyos and T Lange
Stat Methods Med Res published online 23 January 2014
DOI: 10.1177/0962280213520436
Conclusion

- Causal inference has a large potential in event history analysis
- **Time-continuous** view of causality is important. Martingale theory **as useful as ever**
- Local independence graphs à la Schweder allows extension of DAG theory to processes. This yields essential simplifications in understanding marginal structural models and other causal aspects.
- In understanding mediation, the time-continuous aspect is essential. The all too common discretization with respect to time can give wrong results