STRATOS TG7: Causal inference

Causal inference at the intersection of state of the art methods

Bianca De Stavola (LSHTM, UK)
Els Goetghebeur (Gent, B)
Saskia Le Cessie (Leiden, NL)
Erica Moodie (Mc Gill, Ca)
Ingeborg Waernbaum (Umea,S)

Niels Keiding (Copenhagen,D) & Michael Wallace (Mc Gill,Ca)

Banff, July 4, 2016

Motivation

- Causal inference challenging: at conceptual & technical level
- Statistical literature is exploding
  - Several formalisms and schools of thought
  - expanding tool kit
  - questions ever more ambitious (e.g. involving genetics)
- Seemingly 'easy' to do $->$ software
- Applied literature littered with 'suboptimal' causal claims
- Transparency on fundamentals and important 'detail' lacking
- COI: learning, clarifying, bridging, serving, impact, have fun

The TG7 broader plan

- I: 'Target causal effect parameters' of different approaches:
  - their interpretation and practical use/relevance
  - the assumptions involved
  - their overlap and distinction
- II: on estimation under the standard assumptions
  - how it is done (incl. software hints)
  - practical properties of the estimators
  - tricks and treats
- III: What it still means when untestable assumptions fail +
  - Clues on failed assumptions
  - Robustness, sensitivity, and the bias-variance trade off
- IV: Missing data
- V: Some guidance

Links with other topic groups!
descriptives, prediction, missing data...

The TG7 plan - our approach

- work from simple to complex
- from binary trt. to continuous and static or dynamic treatment regimes over time
- from binary over continuous, right censored survival to generally repeated outcomes over time
- from (semi)-parametric to more flexible prediction models
- from (repeated) 'cross-sectional' to longitudinal data set-up, prospective to retrospective designs, ...
- population constant effects and exposures interacting; conditional and average effects
- acknowledging increasing levels of (unmeasured) confounding
- handling missing data

Pointers to tutorials and software implementation

Worked out case studies , simulation studies

From paper(s) to website with links: getting more people involved

Guidance - a compas for practical causal inference

Fundamentals of the methodological approaches

- phrasing the causal question $->$ target estimand
- 'selecting' the observational data:
  - population (Louis & Keiding 2016)
  - exposure (Hernan, 2016)
  - covariates
  - outcome (FUP)
- assumptions to justify causal effect estimators
- chose 'best' analysis, check 'testable assumptions'
- conducting sensitivity analyses for untestable assumptions (Daniel 2013).
- reporting results

Insight is key, simple guidelines and checklist not enough

Among the basic key considerations

Nature of treatment ? Given to whom ?

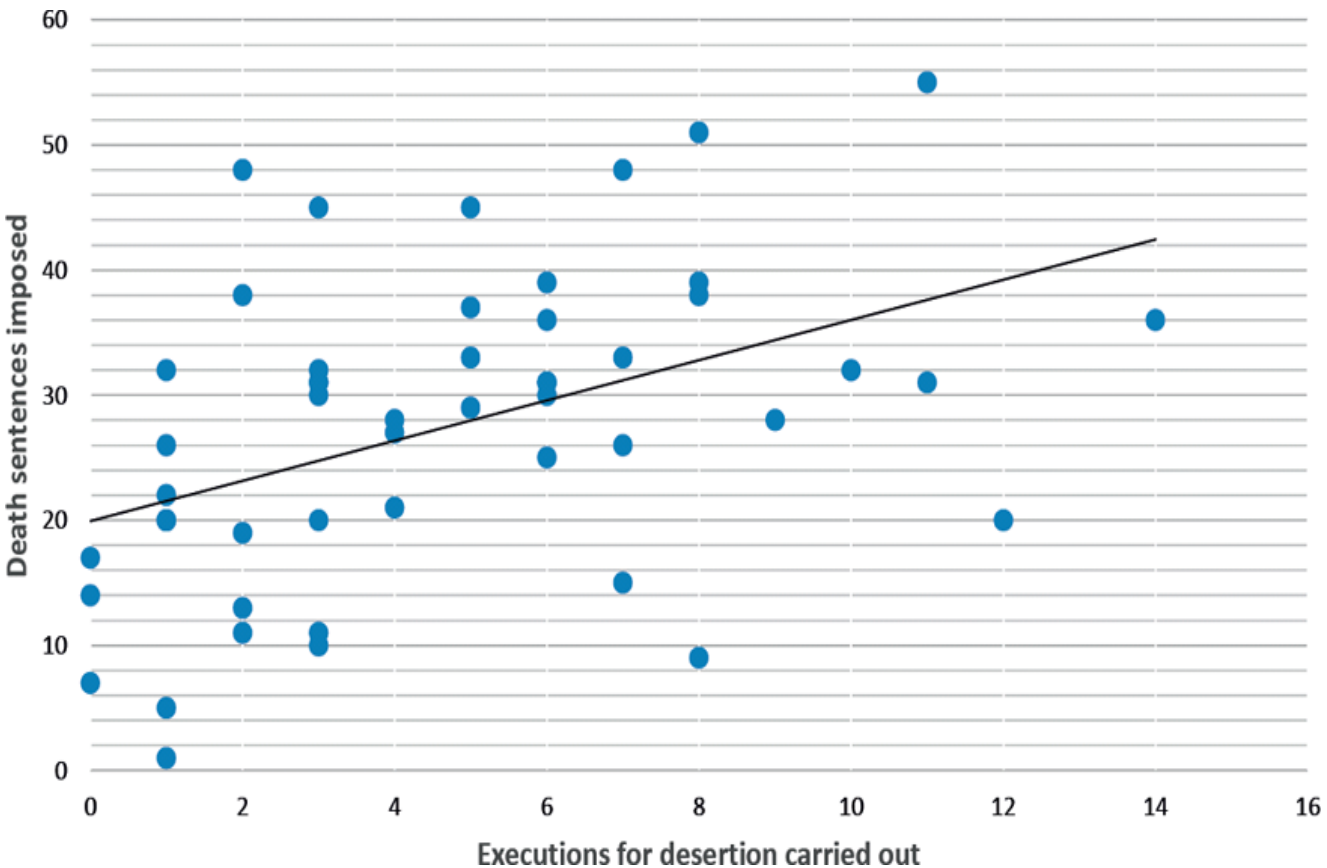
- Internal validity of exposure - outcome association $->$ for prediction purposes
- Internal validity of causal effect estimation: $->$ what did the exposure change for what subgroup? Effect of treatment (if given at all (started), or per protocol, given 'as observed') in
  - The full study population
  - Among the treated
  - Among (full) compliers with treatment ('per protocol')
- External validity: about extrapolating to 'future' interventions/decisions (0)
  - the treatment effect (a specific contrast between 'treatment A' and 'NOT treatment A')
  - the result among the treated (where will they land?)

Practically - paper I on point exposure

The 'mother' paper:

- Classes of causal questions, data structures and assumptions
- Corresponding analysis techniques, their dependence on (untestable) assumptions and available software.
- Tools for transfer to our target audience
  - Website (grant)
  - Infographics
  - Video
  - Simulation 'apps'
  - Case study - with try out options ...

The statistics of tragedy



Significance
Volume 12, Issue 4, pages 10-15, 6 AUG 2015 DOI: 10.1111/j.1740-9713.2015.00838.x
http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2015.00838.x/full#sign838-fig-0003

The Tragedy of Statistics - letter published in February

August 2015 issue of 'Significance', page 10-15

There are of course a number of problems applying statistical analysis to a sample of this limited scope.

The Tragedy of Statistics - letter published in February

August 2015 issue of 'Significance', page 10-15

There are of course a number of problems applying statistical analysis to a sample of this limited scope. The larger the sample the greater the chances that our observations will be accurate and, thus, any hypothesis we develop will gain credibility. However with this limited sampling we can draw some reasonable conclusions.

The Tragedy of Statistics - letter published in February

August 2015 issue of 'Significance', page 10-15

There are of course a number of problems applying statistical analysis to a sample of this limited scope. The larger the sample the greater the chances that our observations will be accurate and, thus, any hypothesis we develop will gain credibility. However with this limited sampling we can draw some reasonable conclusions.

Statistically speaking, the crucial question is whether there can be said to be a negative correlation between executions for desertion and the number of men in a subsequent month who committed this same crime. If the number of men decreases as the number of executions for this crime increases, then it can safely be said that the disciplinary policy of the BAHC worked as a deterrent.

How did involved experts react?

- The journalist editor of Significance
- A famous statistician on the editorial board
- Epidemiologists like Michael Marmott
- The current RSS president

What is the best response?

- Write (letters) about it: head on, do it massively, well written, funny, to the point, make a mark...?
  (by the way: this takes A LOT of time)
- What is our goal, really, how do we make a difference?
  What is our causal effect?
- Be constructive
  Facing the facts versus faking the facts

What is the best response?

- Write (letters) about it: head on, do it massively, well written, funny, to the point, make a mark...?
  (by the way: this takes A LOT of time)
- What is our goal, really, how do we make a difference?
  What is our causal effect?
- Be constructive
  Facing the facts versus faking the facts A standardised risk is just that: a weighted average of conditional associations ('under the model')
  what less ambitious goal can be served?
- Become the new Hans Rosling?
- Big data is hot: big mistakes are too (teach a data mining class)

How come?

What does it take to get the right-ish analysis

- time to think (vs. pressure to publish quickly) about what we have and what we want to have (extrapolation) Need at least a good explanation, e.g. of compliance in the trial)
- subject matter knowledge (assumptions) $->$ collaborators
- insight and guidance in the methods $->$ where TG7 wants to position itself
- tools (software) to perform the analysis ($->$ not the worst part)
- communication about what went in and out of the analysis Challenge: look at published meta-analysis of treatment effects

Less obvious - more impact example: HRT

The Women's Health Initiative: a randomized trial of combined 'estrogen and progestin' in 16,608 women with uterus (JAMA, 2002)

- Primary outcome: CHD
- Motivation: '*A large body of observational studies suggesting a 40% to 50% reduction in risk among users of either estrogen alone or, less frequently, estrogen and progestin*'

Believed to be causal and a multi billion market resulted until...

1. a randomized trial was conducted (WHI, JAMA 2002)
2. a targeted causal analysis of large observational (nurses health) study confirmed the RT results (Hernàn *et al.*, Epidemiology, 2008)

Less obvious - more impact example: HRT

The Women's Health Initiative: a randomized trial of combined 'estrogen and progestin' in 16,608 women with uterus (JAMA, 2002)

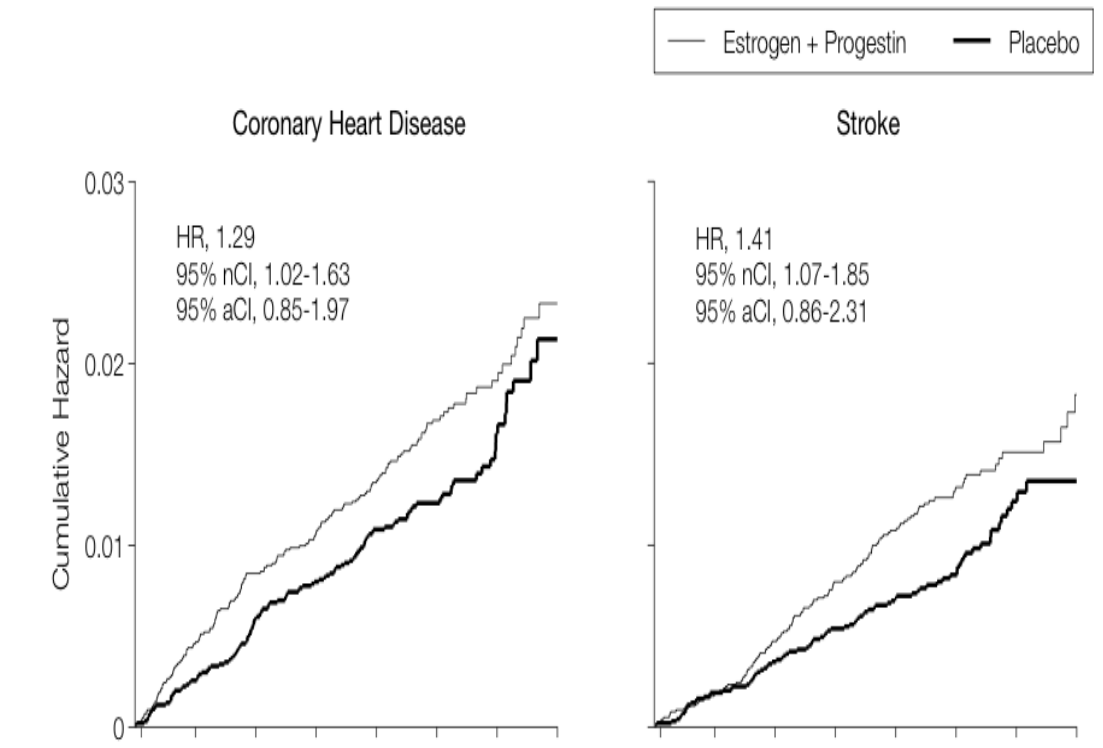
- Primary outcome: CHD
- Motivation: '*A large body of observational studies suggesting a 40% to 50% reduction in risk among users of either estrogen alone or, less frequently, estrogen and progestin*'

Believed to be causal and a multi billion market resulted until...

1. a randomized trial was conducted (WHI, JAMA 2002)
2. a targeted causal analysis of large observational (nurses health) study confirmed the RT results (Hernàn *et al.*, Epidemiology, 2008)

(Also very nice: Danaei et al. 2013, statin (non-)initiators and effect on CHD (CRT emulation in observational study).

Figure 3. Kaplan-Meier Estimates of Cumulative Hazards for Selected Clinical Outcomes

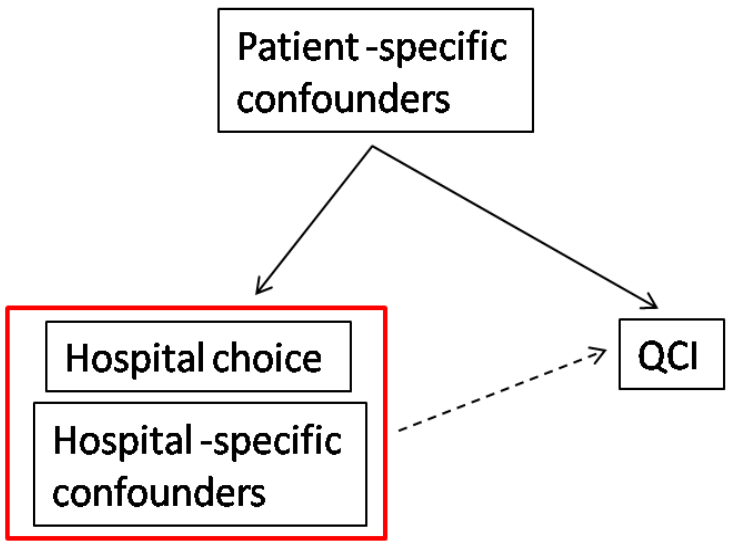


No. at Risk

Estrogen + Progestin	8506	8353	8248	8133	7004	4251	2085	814	8506	8375	8277	8155	7032	4272	2088	814
Placebo	8102	7999	7899	7789	6639	3948	1756	523	8102	8005	7912	7804	6659	3960	1760	524

No unmeasured confounders

General properties: regression and propensity score method

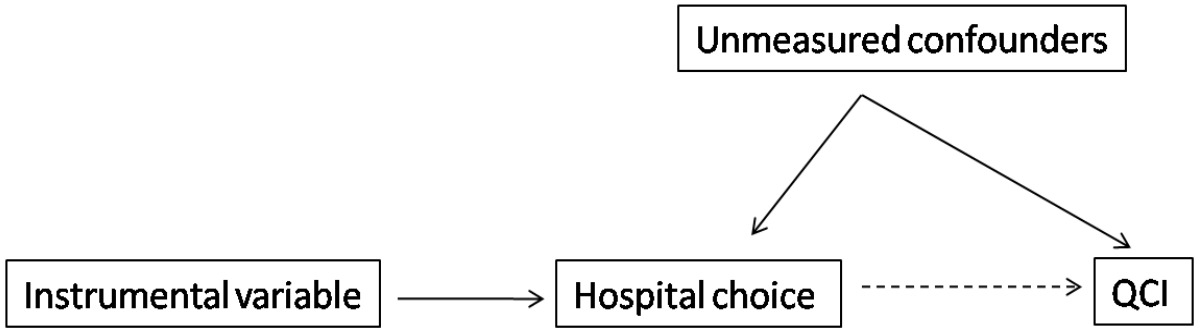


- 'no unmeasured confounders' assumption
- Account for case mix
  - using regression adjustment for outcome (regression)
  - using inverse weighting by propensity to attend specific centre (propensity score method)
- Possibility to explain effect by centre size in second phase

Instrumental variables

General properties: instrumental variables methods

Strongly associated with centre choice (acts as surrogate), but is not associated with the QCI.



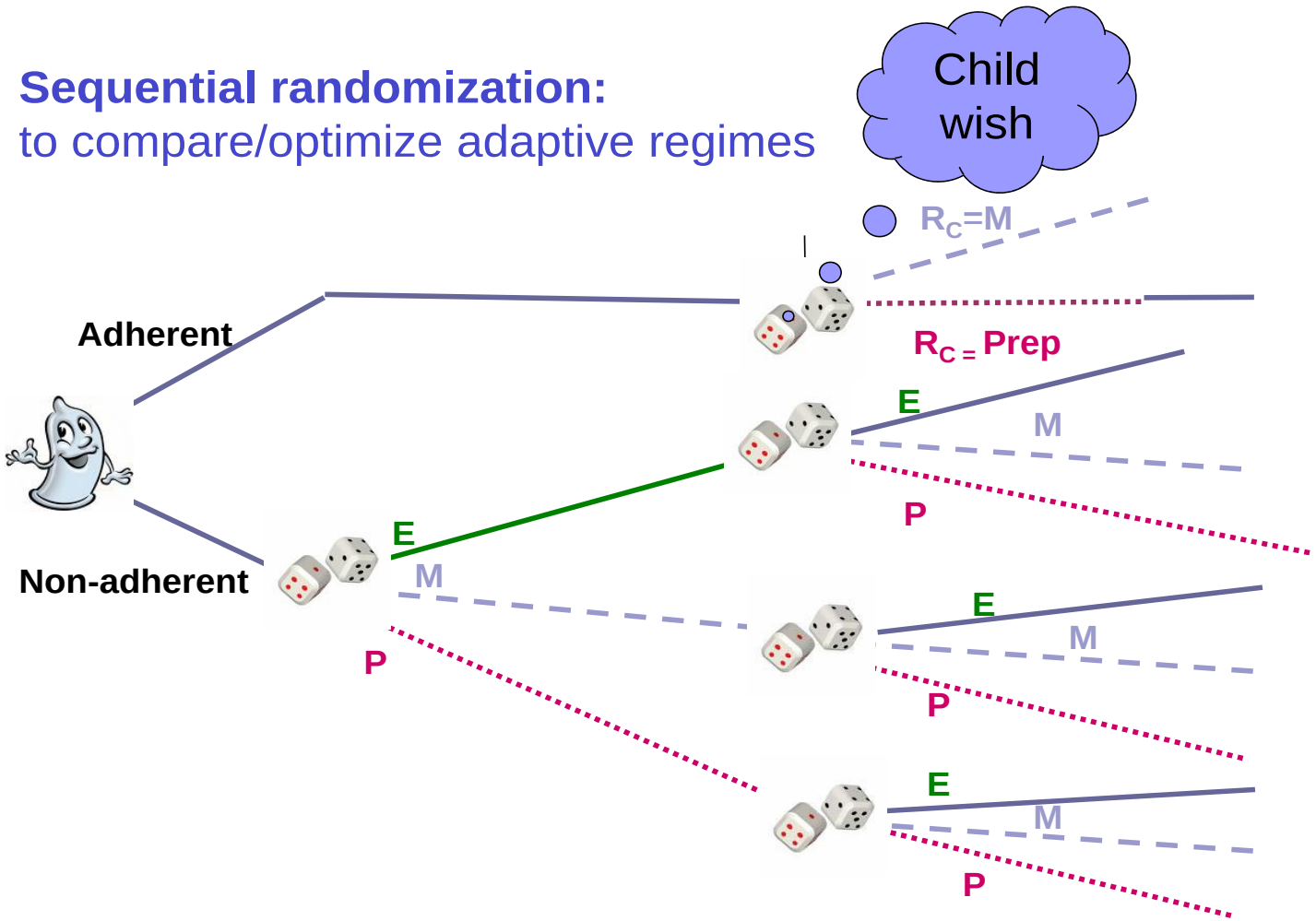
- Account for case mix by relying on pseudo randomization through instrumental variable
- 'unmeasured confounders' allowed

Example: Distance to each centre as IV:

- More likely to choose centre 5km versus centre 200km away.
- Distance to each centre does not tell anything about the QCI.

HIV prevention -design

Sequential randomization:
to compare/optimize adaptive regimes



Outcome regression (Danaei et al. 2013)

$Y(\mathfrak{a}) \coprod A|L \quad \forall \mathfrak{a} \quad \Rightarrow$

$$\{Y|L, A = a\} = \{Y(a)|L, A = a\} \stackrel{d}{=} \{Y(a)|L\}$$

Hence simply regress Y on L in several A-defined strata
to infer the population distribution of $Y(a)$ conditional on L.

regress Y on L in Statin users	$-> f_1(y\|\ell)$
regress Y on L in Non-statin users	$-> f_0(y\|\ell)$

Outcome regression (Danaei et al. 2013)

$Y(\mathfrak{a}) \coprod A|L \quad \forall \mathfrak{a} \quad \Rightarrow$

$$\{Y|L, A = a\} = \{Y(a)|L, A = a\} \stackrel{d}{=} \{Y(a)|L\}$$

Hence simply regress Y on L in several A-defined strata
to infer the population distribution of $Y(a)$ conditional on L.

regress Y on L in Statin users	$-> f_1(y\|\ell)$
regress Y on L in Non-statin users	$-> f_0(y\|\ell)$

Challenges:

- With 'high' dimension of ℓ : confidence in a correct model
- $L-$distribution for (non)treated does not overlap $(\pm)$
  e.g. in the young and fit you may find no statin users
- $E(Y|L, A = 1) - E(Y|L, A = 0) =$
  $E(Y(1)|L) - E(Y(0)|L) = \psi(L)$ i.e. may differ over L

Inverse probability of treatment weighting - QOC

The counterfactual risk $P\{Y(c) = 1\} = E\{Y(c)\}$ in centre c?

Using an outcome working model

$$E(Y|A = c, \mathbf{L}) = m(c, \mathbf{L}; \alpha, \boldsymbol{\beta})$$

and a propensity score working model

$$P(A = c|\mathbf{L}) = h(c, \mathbf{L}; \alpha_c^*, \boldsymbol{\beta}_c^*)$$

$$\hat{E}\{Y(c)\} = \frac{1}{n}\sum_{i=1}^{n} m(c, \mathbf{L}_i; \widehat{\alpha}, \widehat{\boldsymbol{\beta}}) + \frac{1}{n}\sum_{i=1}^{n} \frac{A_{ic}}{h(c, \mathbf{L}_i; \widehat{\alpha}_c^*, \widehat{\boldsymbol{\beta}}_c^*)} \left\{ Y_i - m(c, \mathbf{L}_i; \widehat{\alpha}, \widehat{\boldsymbol{\beta}}) \right\}$$

Special links with TG1: missing data

- Potential outcomes are structurally missing values
  'double missing' for principal strata: also (principal) exposure status no longer observed.
- Techniques for missingness-at-random and treatment level-at-random are closely related:
  both IV based and 'No Unmeasured Confounders' based
- In drug trials treatment non-compliance tends to go had in hand with missing data (e.g. in mental health)
- IV based survival analysis works with backtransformed times, back-transformed non-informative censoring times are no longer non-informative and need special care
- In QOC measurement based on Swedish registers: even limited missingness has an impact on center labelling (MI vs. CC)
  better registration leads to worse labelling when CC analysis is used, less so with MI (based on MAR)

Special links with TG2 and TG6: model selection

Evaluation of prediction error in (20%) validation set: not on target
true target, the causal effect, is hard to reach
One approach fro QOC in Swedish acute stroke register:

- Split dataset 50/50
- Use most informed model (based on all data) to construct average potential outcomes under different exposures in the validation set
- Fit parsimonious models on the training data, select the one that best approaches the target on the validation data
- Note: different covariates selected for prediction & causal effects
  e.g. age of the patient (individual prediction), versus year of entry of the patients (for centre effect estimation)

Special links with TG5: measurement error

- The trade off: more vs. less confounder adjustment
  must be balanced against
  more vs less missing data and measurement error in covariates
- exposure levels actually experienced difficult to measure
  - behavioural: compliance with drug prescription (MEMS), preventive measures (HIV transmission), diet and nutrition,
  - exposure to air pollutants,...
  - big data benefits, EHR

TG8: survival analysis

- Specific causal inference methods have been developed here: e.g. some emphasised the value of AFT methods, additive hazards, ...
- to do: the causal question in a competing risks setting

Something on our resources

- Marie Curie grant MIROR: Methods in research on research
  - One ph.d. topic on 'comparing causal claims across publications"
  - Partners include BMJ, BMC publishers groups, COCHRAN
  - Strong on outreach and communication
    Review of causal claims made in disease specific area, BMJ, ...
- Swedish grant with 5 years part-time support for the website (we have big plans)

Many exciting things to do!

Please come and join us

Lots of skills and work necessary ...