UNIVERSITY OF LEICESTER

Joint models for survival and longitudinal data when the observation process is informative

Alessandro Gasparini[1] Keith R Abrams[1]
Jessica K Barrett[2,3] Michael J Sweeting[1,2] Michael J Crowther[1]

[1] Department of Health Sciences, University of Leicester, Leicester, United Kingdom

[2] Cardiovascular Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

[3] MRC Biostatistics Unit, Cambridge, United Kingdom

7th Survival Analysis for Junior Researchers Conference
Leiden, 24th-26th April 2018

Background

Health care consumption data is being increasingly used in medical research:

- Answering new, more relevant and detailed clinical questions,

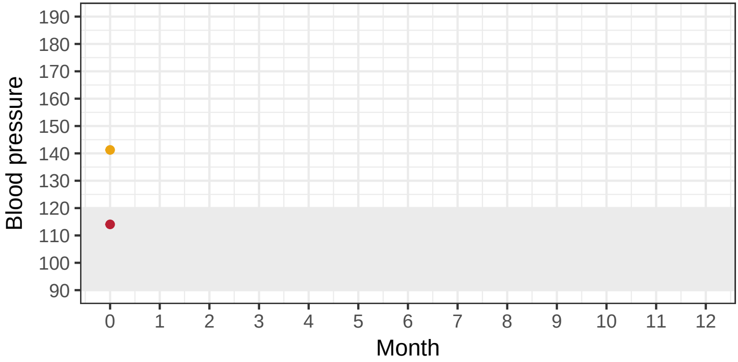
  but...

- New and significant methodological challenges:
  1. Informative censoring;
  2. Informative observation process;
  3. Reporting (REPORT guidelines, Benchimol *et al.*, 2015);
  4. ...

Health care consumption data

In health care records:

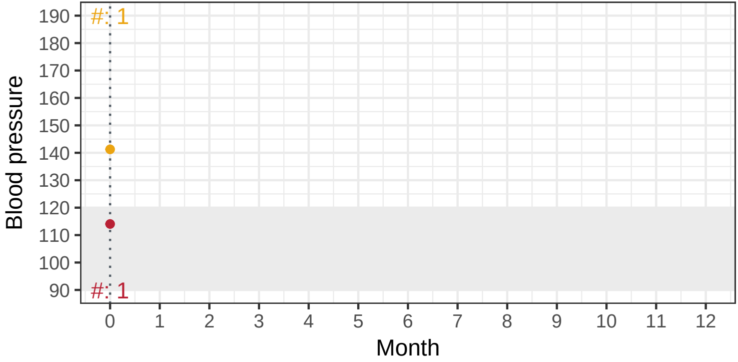
1. Observation times are likely correlated with disease severity;
2. Dropout (censoring) is likely informative.



Health care consumption data

In health care records:

1. Observation times are likely correlated with disease severity;
2. Dropout (censoring) is likely informative.



ag475@leicester.ac.uk

Informative observation process

Common assumption with traditional methods for analysing longitudinal data:

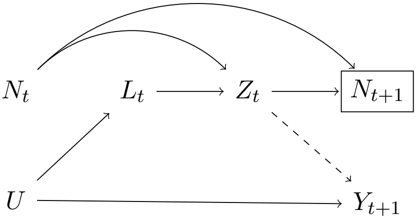
The mechanism that controls the observation time is independent of disease severity

- Joint models for longitudinal-survival data can account for an informative censoring process;
- Research is scarce on whether inference is valid when the observation process is informative.

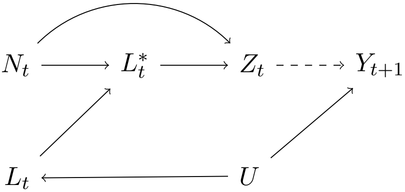
If the observation plan is dynamic, we must account for it in the analysis. Otherwise, two types of bias can arise: selection bias and confounding.

Bias structure

Selection bias:



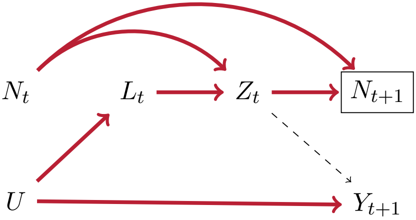
Confounding:



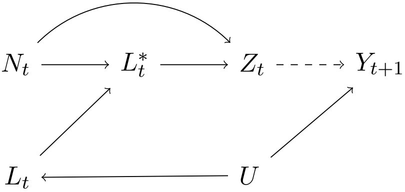
N observation indicator, L covariates, L^* latest measured covariates, Z exposure, Y outcome variable, U unmeasured factors.

Bias structure

Selection bias:



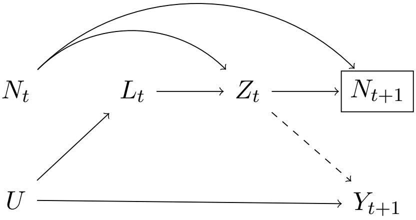
Confounding:



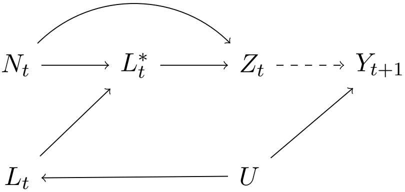
N observation indicator, L covariates, L^* latest measured covariates, Z exposure, Y outcome variable, U unmeasured factors.

Bias structure

Selection bias:



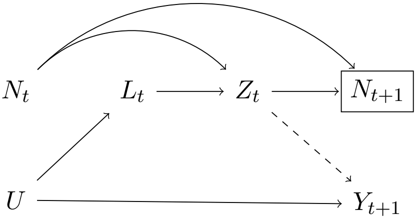
Confounding:



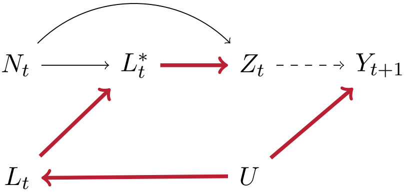
N observation indicator, L covariates, L^* latest measured covariates, Z exposure, Y outcome variable, U unmeasured factors.

Bias structure

Selection bias:



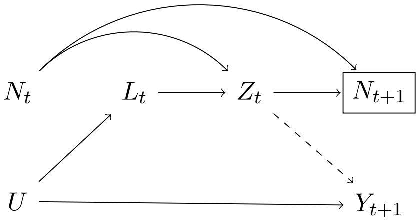
Confounding:



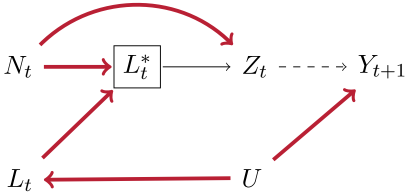
N observation indicator, L covariates, L^* latest measured covariates, Z exposure, Y outcome variable, U unmeasured factors.

Bias structure

Selection bias:



Confounding:



N observation indicator, L covariates, L^* latest measured covariates, Z exposure, Y outcome variable, U unmeasured factors.

State-of-the-art

Some approaches to deal with informative observation times have appeared in the literature. For instance:

- Joint models with random effects (e.g. Liu *et al.*, 2008);
- Methods based on inverse intensity of visit weighting [IIVW] (Robins *et al.*, 1995; Hernán *et al.*, 2009);
- Simple methods such as adjusting for the number of measurements (e.g. Goldstein *et al.*, 2016).

However:

1. there is no real, comprehensive comparison of the performance of different methods in the literature;
2. low awareness of the potential for bias and no guidance (Farzanfar *et al.*, 2017)

A generalised joint model framework

We can fit a generalised multi-equation joint model (Crowther, 2017) to model informative visit times and the longitudinal outcome jointly:

$$r_i = r_0(t)\exp(w_i\beta + u_i) \tag{1}$$

$$y_{ij} | (N_{ij}(t) = 1) = z_{ij}\alpha + \gamma u_i + v_i + \epsilon_{ij} \tag{2}$$

- i and j index individuals and observations, respectively;
- observations of Y_{ij} recorded at each $N_{ij}(t) = 1$;
- z_{ij} and w_i covariate vectors;
- u_i, v_i individual-specific, normally distributed random effects with $E(u) = E(v) = 0$;
- γ association parameter.

A simulation study

Aims: what are the consequences of ignoring the visiting process in practice? How do different methods perform?

True data-generating model (informed by Liu *et al.*, 2008):

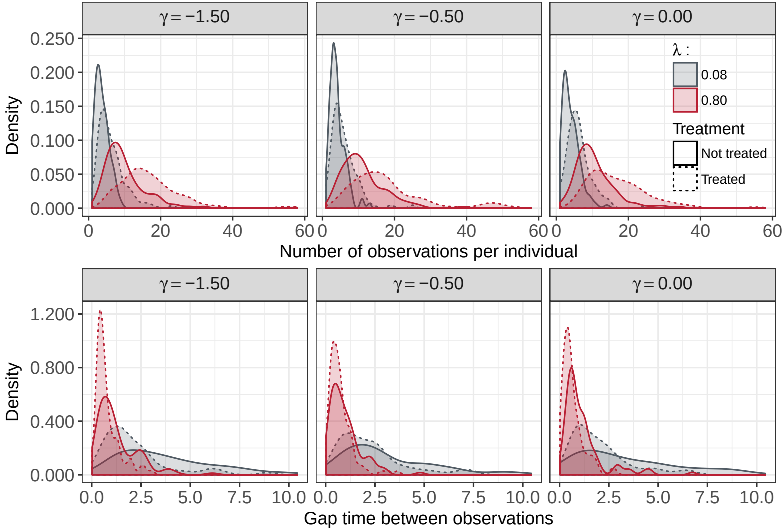
$$r_i = r_0(t)\exp(Z_i\beta + u_i)$$
$$y_{ij}|(dN_{ij}(t) = 1) = \alpha_0 + Z_i\alpha_1 + t_{ij}\alpha_2 + \gamma u_i + v_i + \epsilon_{ij}$$

- binary treatment Z_i;
- $\beta = 1$, $\alpha_0 = 0$, $\alpha_1 = 1$, $\alpha_2 = 0.2$;
- $\sigma^2_u = 1$, $\sigma^2_v = 0.5$, $\sigma^2_\epsilon = 1$;
- $r_0(t)$: Weibull with shape $p = 2$ and scale $\lambda = \{0.08, 0.80\}$;
- $\gamma = \{-1.50, -0.50, 0.00\}$;
- 200 individuals, independent censoring from Unif$(6, 12)$.

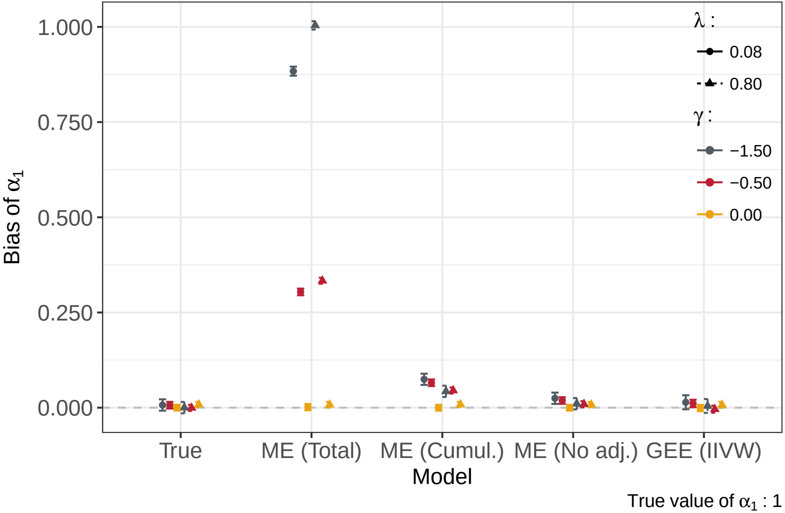
Models included in our comparison

1. True model;
2. A mixed effects model, adjusting for the total number of measurements;
3. A mixed effects model, adjusting for the cumulative number of measurements up to the current time (as a time-varying covariate);
4. A mixed effects model disregarding the observation process;
5. A model fit using generalised estimating equations [GEE] and IIVW (Van Ness *et al.*, 2009).

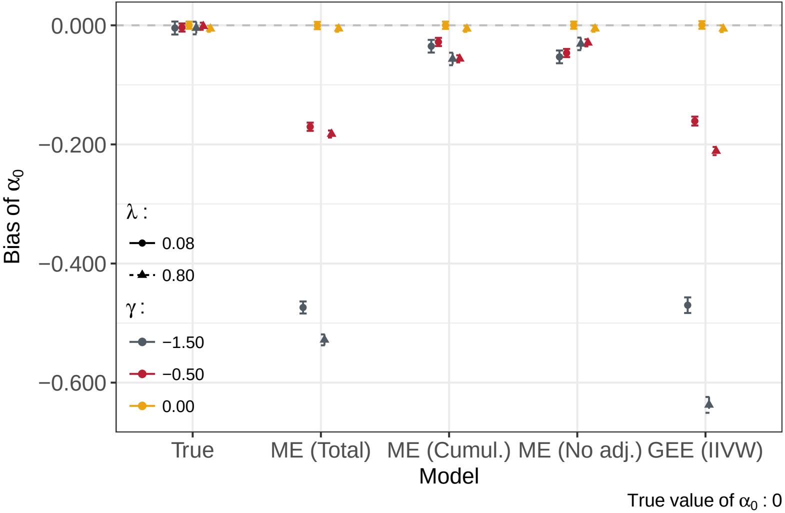
Results: informative observation process



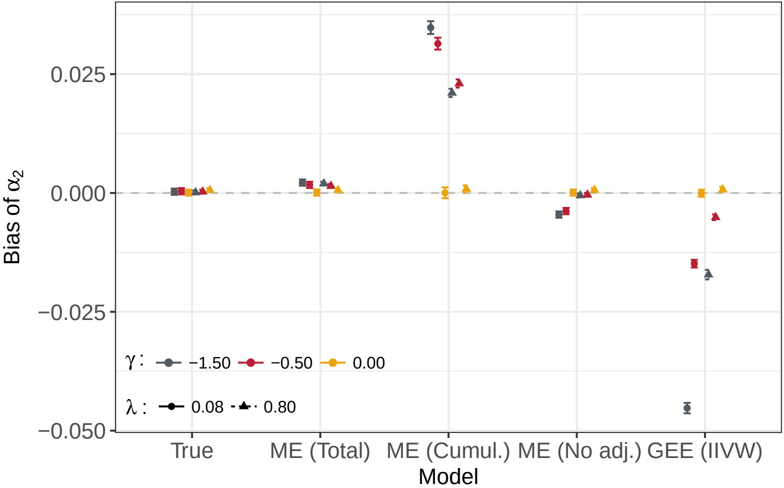
Results: bias of treatment effect



Results: bias of fixed intercept

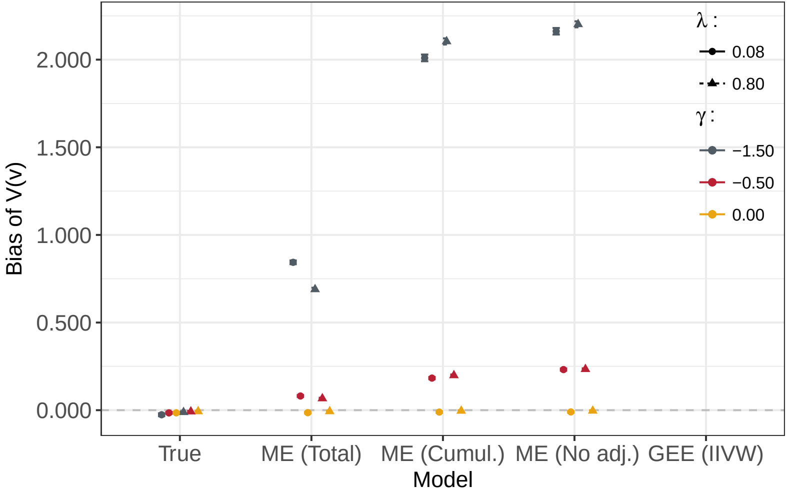


Results: bias of time effect



True value of α_2 : 0.2

Results: bias of variance of random intercept



True value of V(v): 0.5

Conclusions

Take-home messages:

1. Failing to account for a dynamic visiting process yields biased results because of selection bias or confounding;
2. There is a variety of methods that can be utilised to account for an informative visiting process, but they are severely underutilised.

Extension of current work:

- Application to a variety of real data examples;
- Exploring more complex model structures (time-dependent frailties, ...);
- Formalising the joint model in a causal inference framework;
- Additional methods such as multiple outputation (Pullenayegum, 2016).

References

- EI Benchimol, L Smeeth, A Guttmann, K Harron, D Moher, I Petersen, HT Sørensen, E von Elm, SM Langan, RECORD Working Committee (2015). *The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement*. PLoS medicine 12(10):e1001885
- L Liu, X Huang, J O'Quigley (2008). *Analysis of longitudinal data in presence of informative observational times and a dependent terminal event, with application to medical cost data*. Biometrics 64:950-958
- JM Robins, A Rotnitzky, LP Zhao (1995). *Analysis of semiparametric regression models for repeated outcomes in the presence of missing data*. Journal of the American Statistical Association 90(429):106-121
- MA Hernán, M McAdams, N McGrath, E Lanoy, D Costagliola (2009). *Observation plans in longitudinal studies with time-varying treatments*. Statistical Methods in Medical Research 18(1):27-52
- BA Goldstein, NA Bhavsar, M Phelan, MJ Pencina (2016). *Controlling for informed presence bias due to the number of health encounters in an electronic health record*. American Journal of Epidemiology 184(11):847-855
- D Farzanfar, A Abumuamar, J Kim, E Sirotich, Y Wang, EM Pullenayegum (2017). *Longitudinal studies that use data collected as part of usual care risk reporting biased results: a systematic review*. BMC Medical Research Methodology 17(1):133
- MJ Crowther (2017). *Extended multivariate generalised linear and non-linear mixed effects models*. arXiv preprint arXiv:1710.02223, https://arxiv.org/abs/1710.02223
- PH Van Ness, HG Allore, TR Fried, H Lin (2009). *Inverse intensity weighting in generalized linear models as an option for analyzing longitudinal data with triggered observations*. American Journal of Epidemiology 171(1):105-112
- Pullenayegum EM (2016). *Multiple outputation for the analysis of longitudinal data subject to irregular observation*. Statistics in Medicine 35(11):1800-1818