CVD-real and immortal bias

SDCC
http://bendixcarstensen.com/
December 2017
Version 1

Compiled Wednesday 27th December, 2017, 00:59
from: /home/bendix/sdc/proj/daffodil/Suissa.tex

Bendix Carstensen Steno Diabetes Center Copenhagen, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bcar0029@regionh.dk b@bxc.dk
http://BendixCarstensen.com

Contents

1 Introduction and summary

Suissa [1] has raised a critique of the sampling method for the propensity-score matched CVD-real study. The claim is that the study is marred by immortal time bias by excluding parts of the population follow-up from potential inclusion, specifically parts of the follow-up that by definition contains no events.

While it might be slightly confusing to call it immortal time bias, there is little doubt that Suissa's objections are valid. In the following I shall try to explain the rationale behind the critique, the likely order of magnitude of bias as well as possible short and long term remedies for the oversight we made.

2 Immortal time bias?

The normal use of the term "immortal time bias" is for the situation where follow-up a person is included with a covariate value based on a future measurement, such as has for example been ubiquitous in the large literature on diabetes and cancer. This results in follow-up time allocated to a certain covariate value (such as insuin use, based on future exposure), follow-up time which by definition cannot include any event, leading to an under-estimate of the effect of the particular covariate value on the occurrence rate.

In a certain sense the CVD-real is conditioning on the future, namely in that the oGLD persons included are known *not* to switch to SGLT-2i at any time for the remainder of the study. So one could argue that new-user incidents of other drugs *prior* to SGLT-2i initiation should have been possible to include as matches too (and when they were found to match, the corresponding SGLT-2i incident should have been disregarded — or treated as a time-dependent switch of treatment; see below).

Thus the red follow-up time in the illustration in figure 1 of the paper by Suissa [1] is a part of the exposure to oGLD that can never be included as comparator exposure, and it is excluded from this possibility *because* of later SGLT-2i exposure — and by that token with no deaths in it.

It is not correct to argue that also new episodes of oGLD before a randomly chosen oGLD are also excluded; by the selection of a *random* episode among persons without any SGLT-2i exposure any episode is equally likely to contribute, and — this is the crux of Suissa's argument — these episodes are not comparable to the oGLD episodes among persons that later see an SGLT-2i episode. The latter live long enough to move to SGLT-2i, no such thing is known about a person from a randomly chosen episode.

The chosen design of the CVD-real thus selectively excludes oGLD time with no deaths, so the critique by Suissa is indeed correct; event rates on oGLD are overestimated because risk time is missing but no events are missing from the sampling frame.

2.1 Mitigating circumstances

One could argue that the effect is mainly a calendar time effect, and that since we have included index date as well as time since first line treatment in the propensity score matching we have taken the effect pointed out by Suissa into account.

Moreover, as the no. of SGLT-2i users is comparatively small, the absolute bias from this is presumably small.

However, the fact that the bias was not noted, let alone sought quantified, in the original report of the CVD-real requires that it be done, and that Suissa's contribution be acknowledged.

3 Remedies

3.1 Comparable sampling

In order to remedy the bias in the framework of the current study, we would have to allow for inclusion of a random oGLD new-initiation period occurring pre-SGLT-2i in the possible comparison group. Follow-up after such an episode would of course only be until SGLT-2i initiation.

This would be tantamount to selection of a random episode from each person among *all* new-use episodes after 2014. Persons selected with an oGLD episode would then have their follow-up censored at SGLT-2i initiation, and only SGLT-2i episodes selected as random within a person's episodes would be included in the propensity score matching for comparative analysis.

This way we would mimic a complete follow-up of the entire cohort of drug-users (controlling for propensity score). But it would weaken the study because a number of SGLT-2i new-user episodes would be excluded.

It is not entirely clear to me if we could include a randomly selected oGLD episode as oGLD until switch to SGLT-2i, and as SGLT-2i from then on — there would be no matching person at switch time, so my hunch is no, we would have to stop follow-up at switch to SGLT-2i. Which potentially would entail informative censoring...

3.2 Sampling of episodes

Another remedy could be to make the oGLD episodes approximately comparable to the SGLT-2i episodes by requiring that the oGLD-episodes selected had a similar distribution of preceding oGLD episodes as the SGLT-2i episodes. This could for example be done as follows:

- select all SGLT-2i episodes.
- classify each by no. episode since study start, and possibly date and diabetes duration (duration of drug use, that is). Basically, define these three covariates for each SGLT-2i episode.
- derive all oGLD episodes among the non-SGLT-2i exposed persons, and classify them by the same three covariates (no. episode, date, duration of drug use)
- select oGLD episodes to mimic the covariate distribution for the SGLT-2i episodes.

The last step would in principle be a matching where we matched on the 3 defined variables (no. previous episodes, date of episode start and time from first line to episode

start). Thus replacing the propensity score matching by a two-stage procedure where the time-comparability were made in a separate step.

Note that this description is slightly airy, it is not a recipe of how to do it in practice, more elaboration would be needed.

3.3 Time-dependent propensity scores?

Suissa *et al.* recommends "a prevalent new-user design with time conditional propensity scores can be used to avoid this bias"[2]. This is described as propensity score matching in time-strata. The time is not explicitly defined in the paper cited, but from the illustrations in the paper it seems to be either time since first line drug or no. of prescriptions since first line treatment. However, attained age and duration of diabetes or duration of combination therapy could of course also be used as explicit matching variables. However one might just as well compensate for these effects by including them in the model instead of matching on them.

Suissa *et al.*[2] propose to use a time-dependent Cox model for deriving propensity scores, however without explicitly defining either event type, time scale in the Cox model described, let alone which time-dependent variables to include.

It seems that the suggestion is to model the occurrence rate of switch to the index drug, using time-updated clinical variables (in our case this would be current drug exposure and disease history) in the model for SLGT-2i initiation. This way every person would have a propensity score (predicted rate of switching to SGLT-2i) at any time of follow-up. This would be the propensity score to use as covariate in modeling the outcome(s). As far as I can see this requires that the estimated *hazard* of switching be used as propensity score.

Using a parametric model makes this a piece of cake because the predicted rate of switching to the index drug would be available at any timepoint, and hence can be used as propensity score. This way every initiation will have a propensity score associated with it, which should be used as a covariate. But it beats me how the estimated underlying *rate* can be derived from a Cox-model.

We might instead use the HR (that is the linear predictor from the Cox model) alone as propensity score, but that would eliminate the time scale from the propensity score — is that meant to be re-introduced through explicit matching in time-strata? This makes the need for a specification of the timescale in the Cox model so much more urgent.

3.4 The time-dependent Cox-model

Suissa *et al.*[2] mention that the time-dependent Cox-model with 100,000 persons and some 4,000 switches may be infeasible from a computational point of view, because the risk set has to be re-evaluated 4000 times. This is because the Cox-model operates with an overly detailed base-line hazard — in the example mentioned some 4000 (exchangeable!) time-effects. If we instead use an explicit split of time in say 3-month intervals and use a smooth baseline hazard in a Poisson model, the 170,000 GPRD persons followed for some 6-7 years on average would comprise about 1 mill. person-years, and hence after splitting in 3 month intervals some 6 mill. records; each record referring to a 3-month period in the follow-up of a person with the relevant (time-varying) demographic and clinical variables as

covariates and an indicator of initiating the index drug.

And a data set of a few million records can easily be can easily be accommodated for modeling on a modern computer.

4 CVD-real 2do proposal

In the CVD-real study the propensity scores were computed at the initiation dates and the score contained age and time since first-line treatment (and hence the age at first line treatment). This means that the time-effects that Suissa argues to take into account have been partially accounted for through the propensity score. A simple remedy to further control the alleged bias, would be to include current age, current time since first line and, since SGLT-2i is comparatively new on the market, also calendar time of initiation in the modeling of the outcomes.

This can easily be achieved in the CVD-real study by including these variables (evaluated at index dates) in the model for the events of interest.

Hence I think it would be prudent to do the following:

- Acknowledge Suissa's contribution to the improvement of the study by pointing out an overlooked source of bias.
- Include age, time since first line treatment and date (all evaluated at index date) as covariates in the Cox-models for outcomes, in order to get a first handle on the possible size of bias.
- Include the propensity score itself in the current Cox-models for outcomes. This has in itself no bearing on the possible bias pointed out, but it represents a desired calibration of the model.
- Assess the feasibility of analyzing the entire follow-up among all drug-treated diabetes patients in the relevant period with the following covariates for the rates of the events of interest:
  - current age
  - current time since first line drug
  - current time since latest initiation (some of these are the former "index" dates)
  - propensity score as derived from a time-dependent model for SGLT-2i switch.
  - current exposure — note that this variable is not straightforward to define. It could be defined as the latest drug initiated, but some attention should be given to multiple-drug regimens and where to place them.

  The propensity score would have to be computed through an intensity model for the occurrence rate of SGLT-2i initiation ("Cox-model"), using time-updated covariates for all clinical (that is hospitalization) and drug variables.

  Some 200,000 persons would be included from Denmark, all starting, say, 1 December 2014 or at first drug dispense. With follow-up divided in 2 or 3 month periods, and

with an average follow-up of some 2–3 years we would have a dataset of some 2 mill. records, which would be perfectly manageable in a model without further ado. This goes both for the propensity score and for the outcome model.

The final results would only be marginally more elaborate than those already collected from the Nordic countries.

This would presumably be feasible in the update of the study, but hardly worth pursuing in the "old" data.

- In the updated analysis it would further be desirable to set up a proper model for the sequence of major outcomes of interest, HF, other CVD, CVD death and other Death as a multistate model. Setting it up properly would mean that the same questions as previously as well as questions of survival post HF, years lived with HF etc. could be addressed too.

I think we have all the relevant building blocks for these exercises, but hardly the time.

References

[1] S. Suissa. Lower Risk of Death With SGLT2 Inhibitors in Observational Studies: Real or Bias? *Diabetes Care*, 41(1):6–10, Jan 2018.

[2] S. Suissa, E. E. Moodie, and S. Dell'Aniello. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf*, 26(4):459–468, Apr 2017.