Use of Data-Driven Simulations to inform Real-World Survival Analyses

Michal Abrahamowicz[*1]

Marie-Eve Beauchamp[2], Anne-Laure Boulesteix[3],
Tim P. Morris[4], Willi Sauerbrei[5], Jay S. Kaufman[1],
on behalf of the STRATOS Simulation Panel (SP)

[1] **McGill University, Montreal, Canada**
[2] Research Institute of the McGill University Health Centre, Montreal, Canada
[3] LMU Munich, Munich, Germany
[4] MRC Clinical Trials Unit at UCL, UCL, UK
[5] Medical Center - University of Freiburg, Freiburg, Germany

STRATOS INITIATIVE

McGill

Simulation Panel's Mission & Objective of the Current Study

- STRATOS Simulation Panel's Mission:

  Promote more widespread and more accurate use of simulations in both methodological and applied statistical research, through enhancing their:

  (i) Validity (lack of bias, neutrality) [Boulesteix et al., 2018]

  (ii) Reproducibility (accurate reporting, software availability) [Morris et al., 2019]

  (iii) Practical Relevance (plausibility)

- Current Study's Goal (focus on *Applied* research):

  To stimulate use of **data-driven simulations**

  to assess the **impact of specific imperfections in the available data**

  on the results of **real-world time-to-event analyses**

Boulesteix, Binder, Abrahamowicz et al,. *Biom J* 2018.
Morris, White, Crowther, *Stat Med* 2019.

Background: Need to be *more pro-active* when *dealing with imperfections of real-world data*

- Most real-world observational clinical/epi studies recognize (in Discussion) some imperfections in the available data and/or limitations of the study design that may affect the *accuracy* (or even *validity*?) of the results

- Traditionally, this was limited to a lip service, possibly with vague qualitative comments aimed at minimizing the problem**, *e.g.:*

  "Lack of data on disease severity might have affected some of our estimates, but similar problems are common to this area of research."

** Applies also to many papers co-authored by members of our team ☺

Our approach: Quantitative Bias Analysis (QBA) via Data-Driven Simulations

- To make results directly relevant for the data at hand, **we Combine:** Observed Multivariable Real-world data** with Simulating Additional Data items (outcomes and/or covariates) based on carefully defined assumptions

- To simulate time-to-event data **with time-varying exposure**(s)/covariate(s), we rely on the dedicated, validated, "Permutational Algorithm" [Sylvestre & Abrahamowicz, 2008]

** Note this Contrasts with traditional Methods-driven Simulations (in *statistical papers*) that often assess or compare performance of selected methods across a range of (usually hypothetical) plausible data structures

Sylvestre & Abrahamowicz, *Stat Med* 2008.

Implementation of Data-Driven Simulations: Preliminary Steps 1-3

3 Preliminary Steps:

(*see later slides for implementation in 2 illustrative examples)*

1) **Identify relevant Data Imperfection(s)** in your Available Real-World data & (if relevant) carry related Initial Data Analyses

2) Perform relevant, usually **Multivariable, Analyses of the Available Data to get 'naïve' estimates** (Not corrected for the Imperfection(s) identified in step 1) of the relationships between exposure, outcome, and covariates

3) Based on substantive knowledge and/or literature, **Formulate Assumption(s)** regarding how the *available data* can be modified or expanded to create the *oracle dataset* that is corrected for the expected impact of the imperfection identified in step 1 **

(** Several plausible alternative scenarios may be considered here, each implying repeating further steps 4-7)

Implementation of Data-Driven Simulations: Main Steps 4-7

Data Simulations & Analyses (Steps 4-6) to be independently repeated across *m* (e.g., 1000) replications*:

4) **Generate the 'Oracle data'** (**Free of the imperfections** of interest) that combine relevant empirical estimates from step 2 with additional data simulated according to the assumptions from step 3

5) **Modify the 'Oracle data'** from step 4 **to account for imperfection(s)** identified in step 1

6) **Analyze (6a) the 'Oracle' and (6b) the Modified (Imperfect) data** (from steps 4 *and* 5, respectively), *using the same methods*, and **contrast the corresponding results**

7) FINAL Step: **summarize the results** of step 6 across *m* replications and formulate the Conclusions regarding the Impact of the Data Imperfection

* Steps 4-7 must be repeated for each alternative simulated scenario identified in step 3

Example # 1: Impact of omitting cancer stage in a prognostic study of colon cancer mortality

- Goal of the analyses: estimate the independent (adjusted) association of **obstruction of the colon by a tumour ("exposure")** with **all-cause mortality ("outcome")** among patients diagnosed with colon cancer.

- Data source: publicly available dataset from the *survival* R package [Therneau, 2021], with N = 906 colon cancer patients, 175 (19.3%) with the colon obstructed, and 441 deaths during follow-up [Moertel et al., 1995]. Several time-invariant prognostic factors, measured at cancer diagnosis, are available, some associated with both (i) obstruction exposure and (ii) survival, calling for multivariable analyses.

Therneau, *Survival* R package 2021.
Moertel, Fleming, MacDonald et al., *Ann Int Med* 1995.

Example # 1: steps 1 - 3

- *Step 1 (**Imperfection**):* available data do *not* include cancer stage at diagnosis, a powerful predictor of mortality in colon cancer [Quantin et al, 1999], with higher stage likely associated with both obstruction exposure (i.e. potential unmeasured confounder) and some measured covariates

- *Step 2 (**Naïve analyses**):* multivariable Cox proportional hazards (PH) model, with adjustments for measured covariates (but NOT stage), yields **HR = 1.33 for colon obstruction (95% CI: 1.06; 1.68)**

- *Step 3 (**Substantive Assumptions**):* **higher cancer stage** at diagnosis (dichotomized: stage III-IV *versus* I-II) assumed to have **HR = 4.0 for mortality**, and **OR = 1.2 for colon obstruction**, as well as associations with selected measured covariates.
  4 alternative scenarios: with the true **HR = 1.0, 1.3, 1.5 or 2.0** for **colon obstruction.**

Quantin, Abrahamowicz, Moreau et al., *Am J Epidemiol* 1999.

Example # 1: Simulation Methods (*Step 4:* "Oracle data" generation)

- Step 4.1: Across scenarios & 1000 samples use (fixed) real-world data on:
  - (4.1.1) 906 multivariable X vectors (exposure + measured covariates)
  - (4.1.2) Outcomes: Times of 441 events (deaths) + 465 censorings

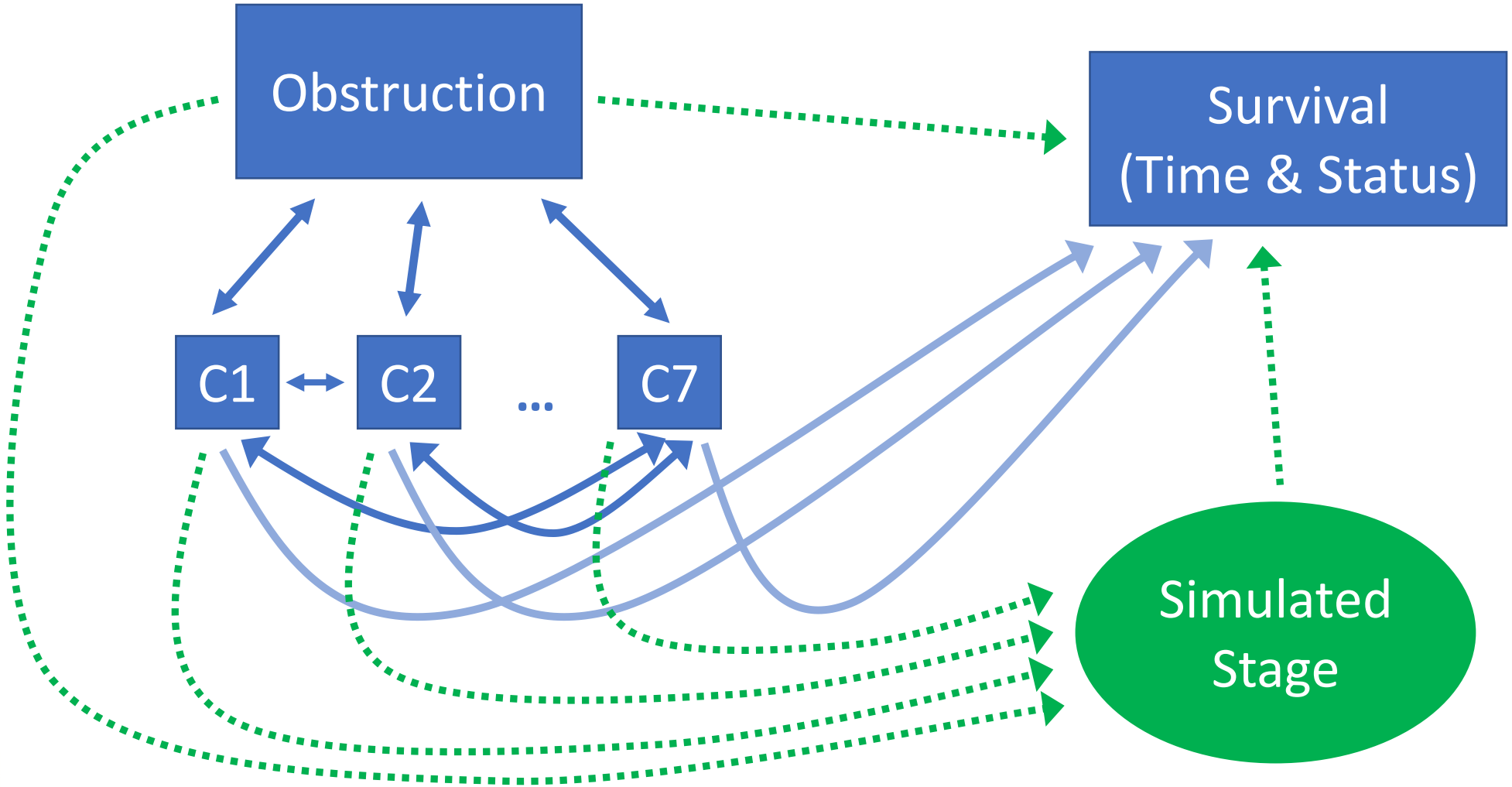
- Step 4.2: **{*Stage* | exposure, covariates} generated independently** in each sample, based on ORs assumed in step 3

- Step 4.3: Use **Permutational Algorithm to assign each of the events or censoring** obs. (with times from 4.1.2) **to one of the 906 'expanded' X vectors** (from 4.1.1 + Stage from 4.2) based on the 'true' PH model, with:

  (i) *Assumed HRs for Stage and Obstruction,* specified in step 3; and

  (ii) For *measured covariates: 'empirical' adjusted HRs estimates* from step 2

Example 1: Observed & Simulated
Data Structure



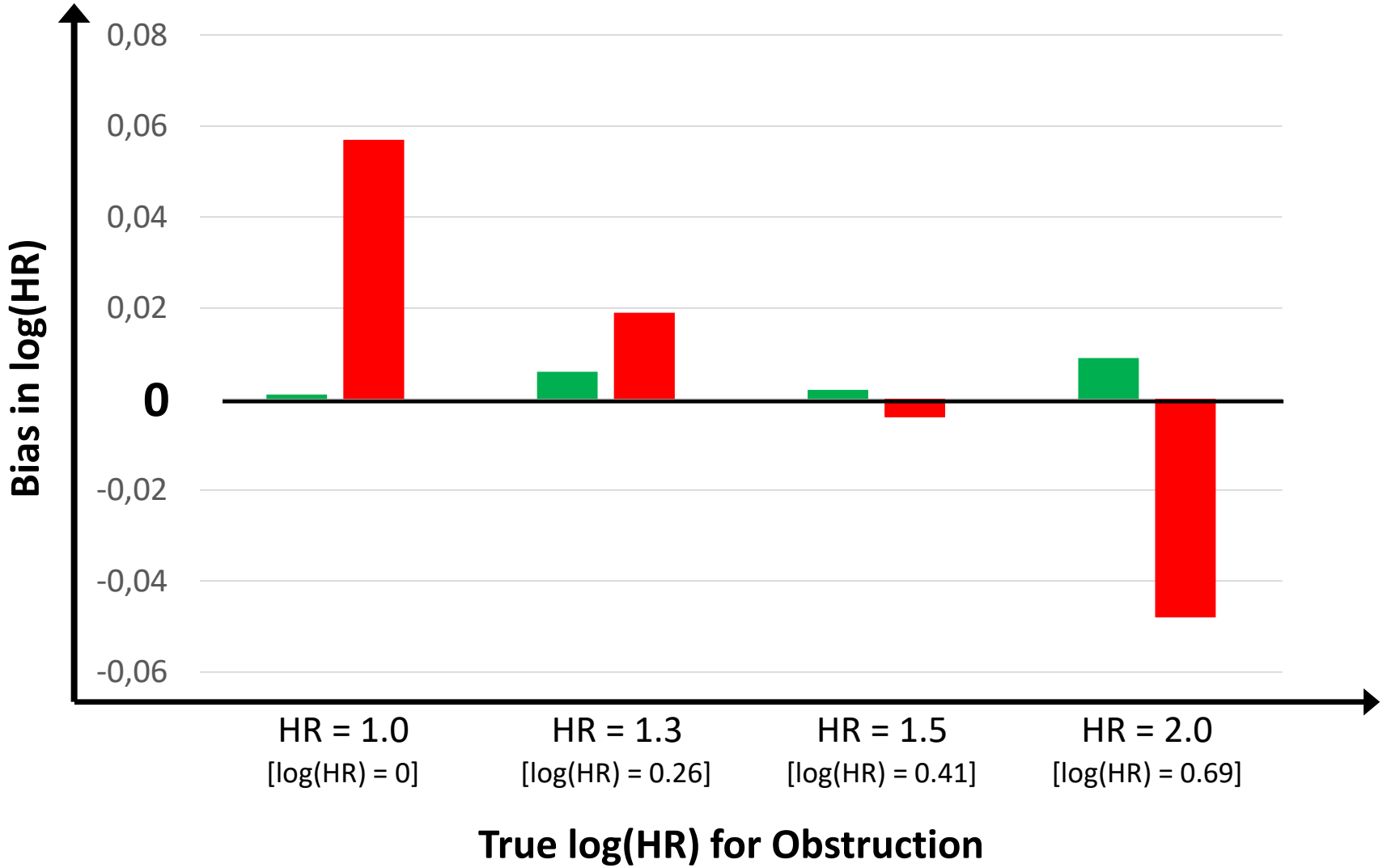
Example # 1: *Steps 5 - 7*

- *Step 5 (Modifying Oracle data to **Insert the Imperfection**):*
  each of *m* samples simulated in step 4 modified by **Deleting "Stage"**

- *Step 6 (**Analyses of (6a) Oracle vs (6b) Imperfect data**):*
  multivariable Cox PH models, with Colon Obstruction & all Measured Covariates, **Only Difference: Stage (6a) Included *vs*. (6b) Stage Excluded**

- *Step 7 (**Summarizing** the results):*
  focus on **BIAS *(Mean of 1000 Estimates – True)* in Adjusted log (HR)** *for Colon Obstruction:* **(6a) *vs*. (6b)**

Example # 1: BIAS in log(HR) for Obstruction as a function of 'true' HR: Oracle vs. Imperfect data



Example # 1: Conclusions

- **Lack of data on Cancer Stage has likely only a minor impact** on the accuracy of the adjusted log(HR) for Colon Obstruction
  (absolute Bias < 0.1, coverage rate of 95% CI: ≥ 90%)

- Expected Bias varies depending on the strength of the (assumed) true association**:
  - (i) Slight over-estimation of null or weak effects (1 ≤ HR ≤ 1.3) *versus*
  - (ii) Slight under-estimation of stronger effects (HR ≥ 1.5)

** Due to a **Combination** of (i) **Unmeasured Confounding** (OR = 1.2 for Stage-Obstruction) ***vs.*** (ii) **Non-Collapsibility** (HR = 4.0 for omitted Stage)

Example # 2: **Association of a Time-Varying exposure with an (imprecisely timed) Interval-Censored event**

- Goal of the analyses: estimate the association of **recent benzodiazepine use** with **cognitive impairment**

- Data source: ***synthetic data*** **based on real-world time-varying patterns of benzodiazepine use** [Bartlett et al., 2004], with N= 1250 new benzodiazepine users generating **285 (23%) events of cognitive impairment** during up to 3 years of follow-up. 2 measured time-invariant covariates: sex and age. **Binary Time-Varying Exposure (TVE) = Any Benzodiazepine use in the last 2 weeks**.

Bartlett, Abrahamowicz, Tamblyn, et al., *Pharmacoepidemiol Drug Saf* 2004.

Example # 2: Steps 1 - 3

- *Step 1 (**Imperfection**):* an event of **cognitive impairment is detected** only at the time of the first clinic visit after its actual occurrence, so the **actual event times remain unknown**, resulting in **Interval-Censored events (see Next Slide for Implications for TVE analyses)**

- *Step 2 (**Naïve analyses**):* 2 Cox PH models (adjusted for age & sex), with ***alternative Event Times Imputation:***

  (i) @ **END** of the Interval (visit when event 'detected'): HR(TVE) = **1.2**; (95%CI: **0.87**-1.68) *vs.*

  (ii) @ **MID-Point** of the Interval (between 2 adjacent visits): HR(TVE) = **1.47**; (95%CI: 1.09-2.00)

- *Step 3 (**Substantive Assumptions**):* **True (UN-known) event equally likely to occur at any time within the between-visit interval** at the end of which it was detected.

  4 alternative scenarios: with HR = 1.0, 1.5, 2.0 or 2.5 for TVE (recent benzodiazepine use)

Impact of inaccurate timing of interval-censored events on the associated 'current' values of time-varying exposure

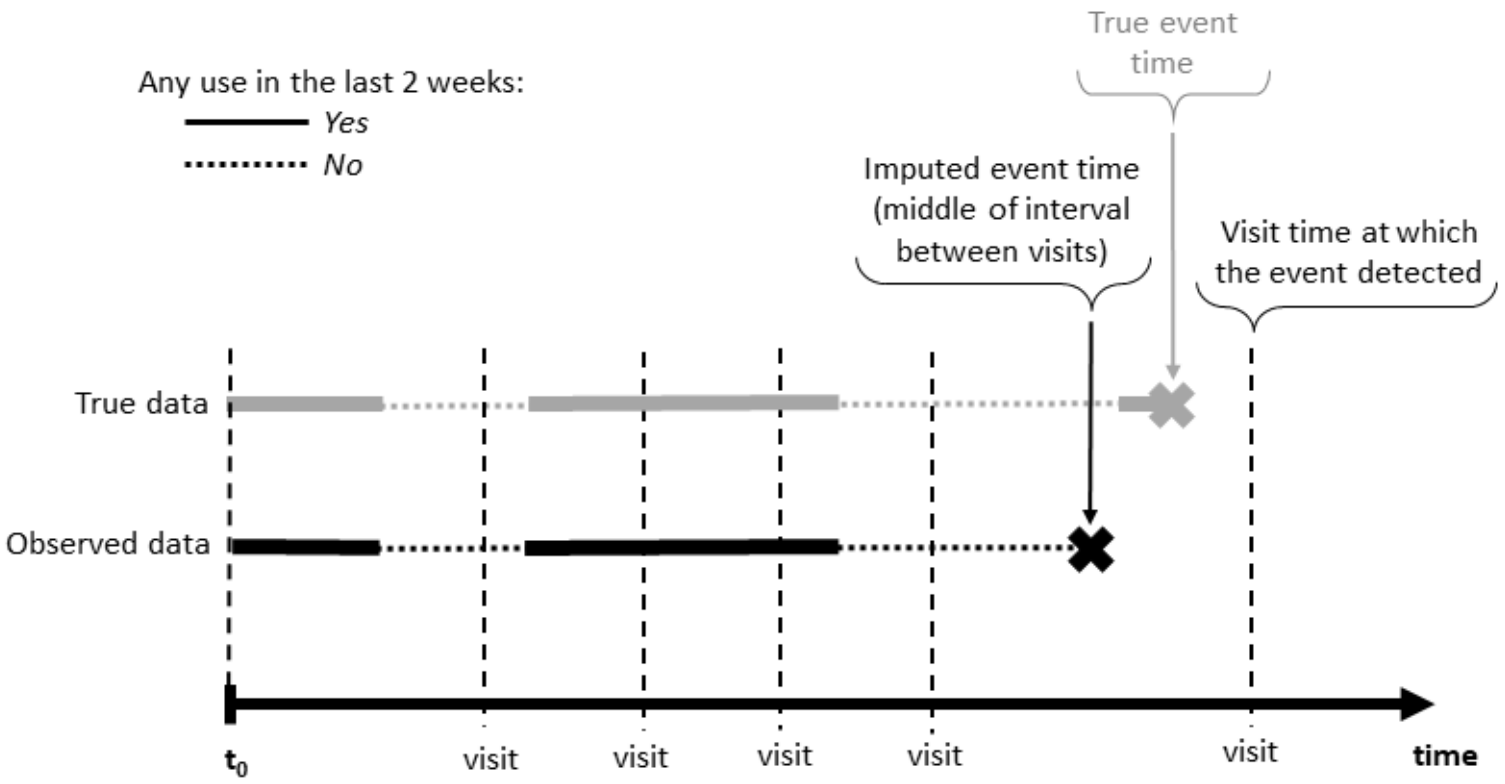


Illustration of the impact of inaccurate timing of interval-censored events for a hypothetical subject: the time-varying exposure metric "any use in the last 2 weeks" value differ between the true event time (exposure = yes) and the imputed event time (exposure = no) at the middle of the intervals between the visits when the event was detected and the preceding visit.

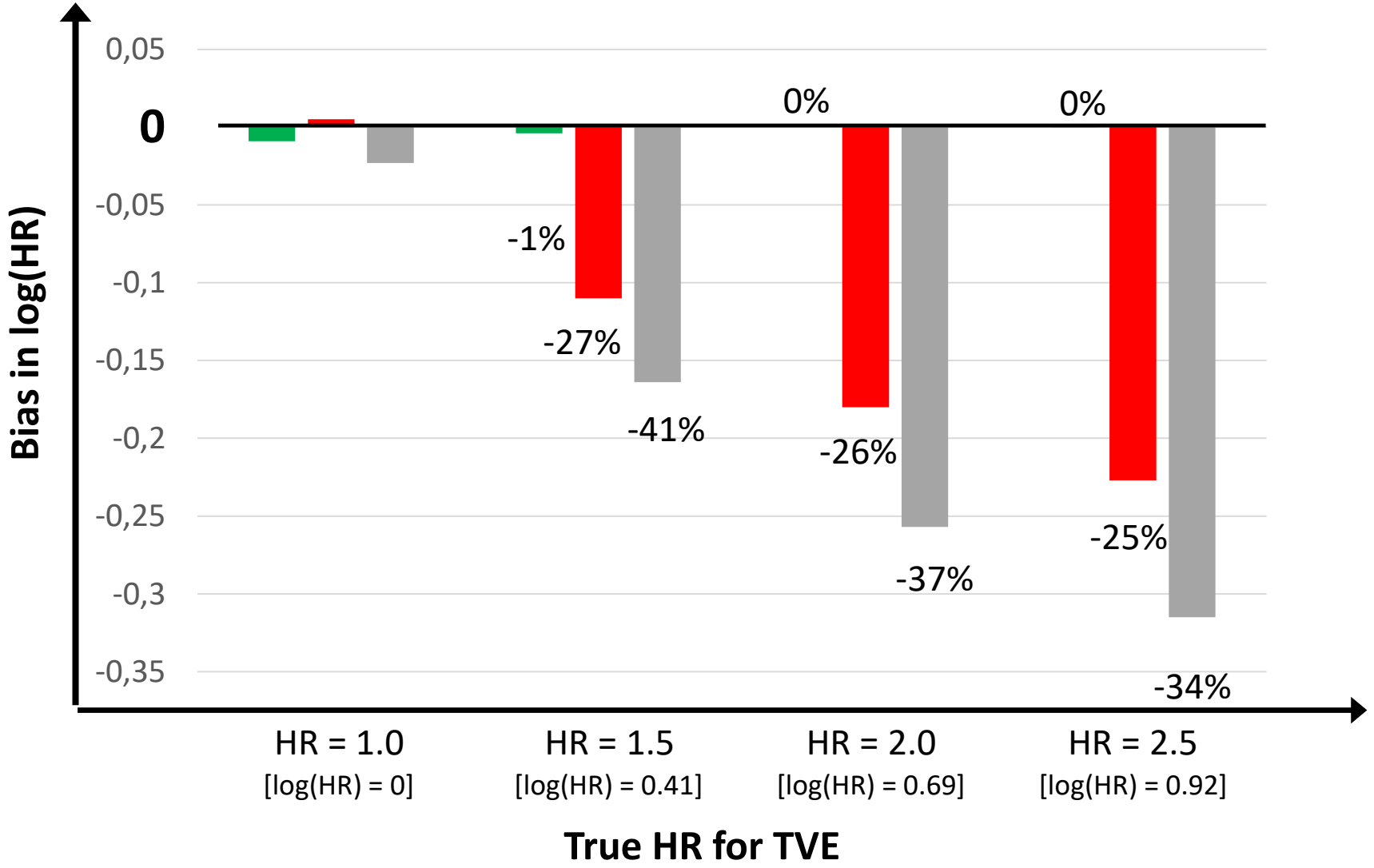
Example # 2: Steps 4 - 6

- *Step 4:* (**"Oracle data" generation):**
  - (4.1) Across scenarios & 1000 samples use (fixed) observed data on: 1250 TVE time-vectors (daily benzo use) with corresponding age & sex
  - (4.2) **"True" Time of event** i = 1,...,285 generated (independently for each of m samples) from ***Uniform U ~ [$t_{i,(j-1)}$; $t_{i,j}$] over the interval*** between the earlier visit $t_{i,(j-1)}$ and visit $t_{i,j}$ when it was detected
  - (4.3) Use **Permutational Algorithm to assign each of the events** obs. (with times from 4.2) **to one of the TVE vectors** (from 4.1) based on the 'true' PH model, with: (i) assumed HR for TVE (from step 3) and (ii) empirical HR estimates for age and sex (from step 2)

- *Step 5: (**Inserting the Imperfection**):* **Exact ('true') event times were Deleted** and only the times of the visits when events were detected were reported

- *Step 6: (**Analyses**): 3* **multivariable Cox models: (6a) Oracle data (True event times) *vs.* (6b) Event Times Imputed (Imperfect data) at: *(6b1) End (detection visit)* *or (6b2) Mid-point* *of the interval* [$t_{i,(j-1)}$; $t_{i,j}$]**

Example # 2: BIAS in log(HR) for TVE as a function of 'true' HR: Oracle vs. Imputation @: MID vs END



Example # 2:
Summary of Results & Conclusions

- **(i) Imprecise Timing of the events** (transient Cognitive Impairment) **induces considerable Bias to the Null in the estimated HR for the Time-Varying Exposure** (recent use of Benzodiazepines)

- **(ii) Bias is systematically stronger for Imputing the events at the End (~ 35-40% relative bias) than at the Mid-point (~ 25% relative bias) ** of the interval between the adjacent visits**

  (** Also, Root Mean Squared Error (RMSE) of End-imputed estimates is 20%-30% higher than for Mid-point-imputed estimates

- (iii) Given (i) & (ii), the **'naïve' estimate based on Mid-point Imputation** of event times [**HR = 1.47** (95%CI: 1.09-2.00)] **provides a solid evidence of Risk Increase associated with a recent Benzodiazepines use but likely Underestimates its strength !**

Conclusions

- **Carefully designed Data-Driven Simulations** can provide valuable insights regarding the **expected impact of a specific Data Imperfection** or Design Limitation on the results and conclusions of a **particular Real-World study**

- **<u>Our methods extend the QBA toolbox to address complexities of:</u>**
  - *Multivariable* data structures
  - *Time-to-Event* (Survival) analysis
  - *Time-Varying Exposures*/Covariates

  but further real-world applications are necessary to fully assess their practical usefulness/potential…

Thank you!

Michal.Abrahamowicz@McGill.ca

References

- **Abrahamowicz M, Beauchamp ME, Boulesteix AL, Morris TP, Sauerbrei W, Kaufman JS, on behalf of the STRATOS Simulation Panel (SP). Data-driven simulations to assess the impact of study imperfections in time-to-event analyses. Under Review for *American Journal of Epidemiology, (March* 2023).**
- Banack HR, Hayes-Larson E, Mayeda ER. Monte Carlo Simulation Approaches for Quantitative Bias Analysis: A Tutorial. *Epidemiologic Reviews* 2021; 43(1):106-117.
- Barberio J, Ahern TP, MacLehose RF, Collin LJ, Cronin-Fenton DP, Damkier P, Sorensen HT, Lash TL. Assessing techniques for quantifying the impact of bias due to an unmeasured confounder: an applied example. *Clinical Epidemiology* 2021; 13:627-635.
- Bartlett G, Abrahamowicz M, Tamblyn R, Grad R, Capek R, du Berger R. Longitudinal patterns of new benzodiazepine use in the elderly. *Pharmacoepidemiology and Drug Safety* 2004; 13(10):669–682.
- Boulesteix AL, Binder H, Abrahamowicz M**,** Sauerbrei W; Simulation Panel of the STRATOS Initiative. On the necessity and design of studies comparing statistical methods. *Biometrical Journal* 2018; 60(1):216-218.
- Lash TL, Fink AK, Fox MP. Unmeasured and unknown confounders. In: *Applying Quantitative Bias Analysis to Epidemiologic Data*. Springer; 2009:59-78.
- Moertel CG, Fleming TR, MacDonald JS, Haller DG, Laurie JA, Tangen CM, Ungerleider JS, Emerson WA, Tormey DC, Glick JH, Veeder MH, Maillard JA. Fluorouracil plus Levamisole as an effective adjuvant therapy after resection of stage II colon carcinoma: a final report. *Annals of Internal Medicine* 1995; 122(5):321-326.
- Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in Medicine* 2019; 38:2074-2102.
- Quantin C, Abrahamowicz M, Moreau T, Bartlett-Esquilant G, MacKenzie T, Tazi MA, …, Faivre J. Variation over time of the effects of prognostic factors in a population based study of colon cancer: Comparison of statistical models*. American Journal of Epidemiology* 1999; 150:1188–1200.
- Sylvestre, M. P. & Abrahamowicz, M. Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine* 2008; 27:2618–2634.
- Therneau T. A Package for Survival Analysis in R. R package version 3.2-13, 2021. https://CRAN.R-project.org/package=survival.
- VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine* 2017; 167(4):268-274.

Outline

- Background / Rationale
- Overview of the proposed Approach to Data-driven Simulations
- 2 Real-World Illustrations:
    1) *Imprecise timing of (**interval-censored) events*** associated with a ***time-varying exposure***
    2) *Omitting* an important risk factor (*potential Confounder*)
- Conclusions

Limitations of existing QBA approaches

- (A) Neither approach accounts fully for the complex Multivariable data structure actually encountered in a given real-world dataset (relationships of different covariates with (i) each other, (ii) exposure, and (iii) outcome)
- (B) Not well developed for Time-to-Event analysis (for which the outcome is often dichotomized, i.e. *Event Times are ignored*) [e.g., Barberio et al., 2021]
- (C) Not clear if/how to handle Time-Varying Exposures (or time-varying covariates) ?

Barberio, Ahern, MacLehose et al., *Clin Epidemiol* 2021.

Simulation results (Example 1): Impact of not adjusting for cancer stage on adjusted log(HR) estimates for colon obstruction

Scenario #	True HR obstruction	True OR obstruction ↔ stage	True HR stage	Performance measures for estimated obstruction log(HR)	Oracle model WITH stage	Original model withOUT stage
1	1.0	1.2	4.0	Bias	0.010	0.064
				SD of estimates	0.124	0.126
				RMSE	0.125	0.142
				Coverage rate 95% CI	0.947	0.900
				Type I error rate (%)	5.3	10.0
2	1.3	1.2	4.0	Bias (relative bias)	0.000 (0%)	0.016 (6.0%)
				SD of estimates	0.116	0.119
				RMSE	0.116	0.120
				Coverage rate 95%CI	0.953	0.951
				Power (%)	60.4	66.0
3	1.5	1.2	4.0	Bias (relative bias)	0.004 (0.9%)	-0.009 (-2.3%)
				SD of estimates	0.115	0.116
				RMSE	0.115	0.116
				Coverage rate 95%CI	0.963	0.950
				Power (%)	94.2	92.6
4	2.0	1.2	4.0	Bias (relative bias)	0.011 (1.5%)	-0.044 (-6.4%)
				SD of estimates	0.118	0.121
				RMSE	0.118	0.129
				Coverage rate 95%CI	0.940	0.921
				Power (%)	100	100

Background: Main existing approaches for Quantitative Bias Analysis (QBA)

- Recent studies incrementally rely on QBA to get a *Quantitative assessment of the potential impact of selected common data imperfections* (e.g., unmeasured confounder or exposure measurement error) [Lash et al., 2009]

- **2 main existing Alternative QBA approaches [Banack et al., 2021]:**

  1) *Analytical correction formulas* for selected, relatively simple analyses, including e.g. E-values for unmeasured confounding [Vanderweele & Ding, 2017], OR

  2) *Simulating Synthetic data*, with data structure generally similar to the real-world data used in a given empirical study

Lash, Fink, Fox, Springer 2009.
Banack, Hayes-Larson, Mayeda, *Epidemiol Rev* 2021.
Vanderweele, Ding, *Ann Int Med* 2017.

Simulation results (Example 1): continued

Scenario #	True HR obstruction	True OR obstruction ↔ stage	True HR stage	Performance measures for estimated obstruction log(HR)	Oracle model WITH stage	Original model withOUT stage
5	1.3	1.0	4.0	Bias	0.004 (1.5%)	-0.034 (-13.0%)
				SD of estimates	0.120	0.123
				RMSE	0.120	0.127
				Coverage rate 95% CI	0.953	0.948
				Type I error rate (%)	60.7	48.3
6	1.0	2.0	4.0	Bias (relative bias)	-0.006	0.214
				SD of estimates	0.120	0.116
				RMSE	0.121	0.243
				Coverage rate 95%CI	0.947	0.565
				Power (%)	5.3	43.5
7	1.3	2.0	4.0	Bias (relative bias)	0.002 (0.8%)	0.177 (67.7%)
				SD of estimates	0.119	0.120
				RMSE	0.119	0.214
				Coverage rate 95%CI	0.944	0.638
				Power (%)	60.4	95.9

Simulation results (Example 2): Comparison of estimates for time-varying recent benzodiazepine use

True HR exposure	Performance measures	Model 1: Oracle	Model 2: Events at MID intervals	Model 3: Events at END intervals
1.0	Bias	-0.011	-0.004	-0.025
	SD of estimates [ratio END/MID]	0.187	0.182	0.184 [1.01]
	RMSE [ratio END/MID]	0.188	0.182	0.185 [1.02]
	Coverage rate 95%CI	0.943	0.955	0.955
	% samples MID closer to TRUTH than END	48.6%		
1.5	Bias (relative bias, %) [ratio bias END/MID]	0.000 (-0.1%)	-0.105 (-26.0%)	-0.161 (-39.6%) [1.52]
	SD of estimates [ratio END/MID]	0.158	0.160	0.165 [1.03]
	RMSE [ratio END/MID]	0.158	0.192	0.230 [1.20]
	Coverage rate 95%CI	0.955	0.922	0.866
	% samples MID closer to TRUTH than END	59.5%		
2.0	Bias (relative bias, %) [ratio bias END/MID]	-0.009 (-1.3%)	-0.186 (-26.9%)	-0.252 (-36.4%) [1.35]
	SD of estimates [ratio END/MID]	0.150	0.151	0.154 [1.02]
	RMSE [ratio END/MID]	0.150	0.240	0.295 [1.23]
	Coverage rate 95%CI	0.948	0.779	0.672
	% samples MID closer to TRUTH than END	67.0%		
2.5	Bias (relative bias, %) [ratio bias END/MID]	-0.004 (-0.4%)	-0.230 (-25.1%)	-0.314 (-34.3%) [1.36]
	SD of estimates [ratio END/MID]	0.139	0.140	0.147 [1.05]
	RMSE [ratio END/MID]	0.139	0.269	0.347 [1.29]
	Coverage rate 95%CI	0.953	0.676	0.454
	% samples MID closer to TRUTH than END	72.5%		

Example # 1: Impact of omitting cancer stage in a prognostic study of colon cancer mortality

- Goal of the analyses: estimating independent (adjusted) association of **obstruction of the colon by a tumour ('exposure')** with **all-cause mortality ('outcome')** among patients diagnosed with colon cancer
- Data source: publicly available dataset from the *survival* R package [1, Therneau, 2021], with N=906 colon cancer patients, 175 (19.3%) with the colon obstructed, and 441 deaths during the follow-up [2, Moertel et al., 1995]. Several time-invariant prognostic factors, measured at cancer diagnosis, are available [1,2], some associated with both (i) obstruction exposure and (ii) survival, calling for multivariable analyses.
- *Step 1 (**Imperfection**):* the available data do *not* include cancer stage at diagnosis, a powerful predictor of mortality in colorectal cancer [Quantin, 1999], with higher stage likely associated with both obstruction exposure (i.e. potential unmeasured confounder) and some measured covariates
- *Step 2 (**Naïve analyses**):* multivariable Cox proportional hazards (PH) model, with adjustments for measured covariates (but NOT stage), yields **HR=1.33 for colon obstruction (95% CI:1.06; 1.68)**
- *Step 3 (**Substantive Assumptions**): 7 alternative scenarios:* **higher cancer stage** at diagnosis (dichotomized: stage III-IV *versus* I-II) assumed to have: **HR=4.0 for mortality**, and **OR=1.2 for colon obstruction** (*modified in some scenarios*), as well as associations with selected measured covariates. Across the scenarios, the true HR for the exposure **(colon obstruction) varied (HR= 1.0, 1.3, 1.5 or 2.0).**

Therneau, R package 2021.
Moertel, Fleming, MacDonald et al., *Ann Int Med* 1995.
Quantin, Abrahamowicz, Moreau et al., *Am J Epidemiol* 1999.