Immortal bias - you live longer if you cannot die!

Kaspar Rufibach
Methods, Collaboration & Outreach Group, Department of Biostatistics, Roche Basel
14th Basel Modeling & Simulation Seminar, 10th September 2018

Generated 2018-09-04 at 21:15:21.

Ackowledgments

Inspired by an earlier presentation by the **Methods, Collaboration & Outreach** (MCO) group within Roche Biostatistics.

Feedback on earlier versions by

- **Marcel Wolbers**,
- **Nicolas Frey**, especially pointing out the Herceptin example,

gratefully acknowledged.

Hypothetical introductory example

Hypothetical example:

- Patients admitted to intensive care unit, ICU = time origin.
- Goal: Assess mortality of new "treatment": a **cup of tea on day 15**, compared to "no treatment".

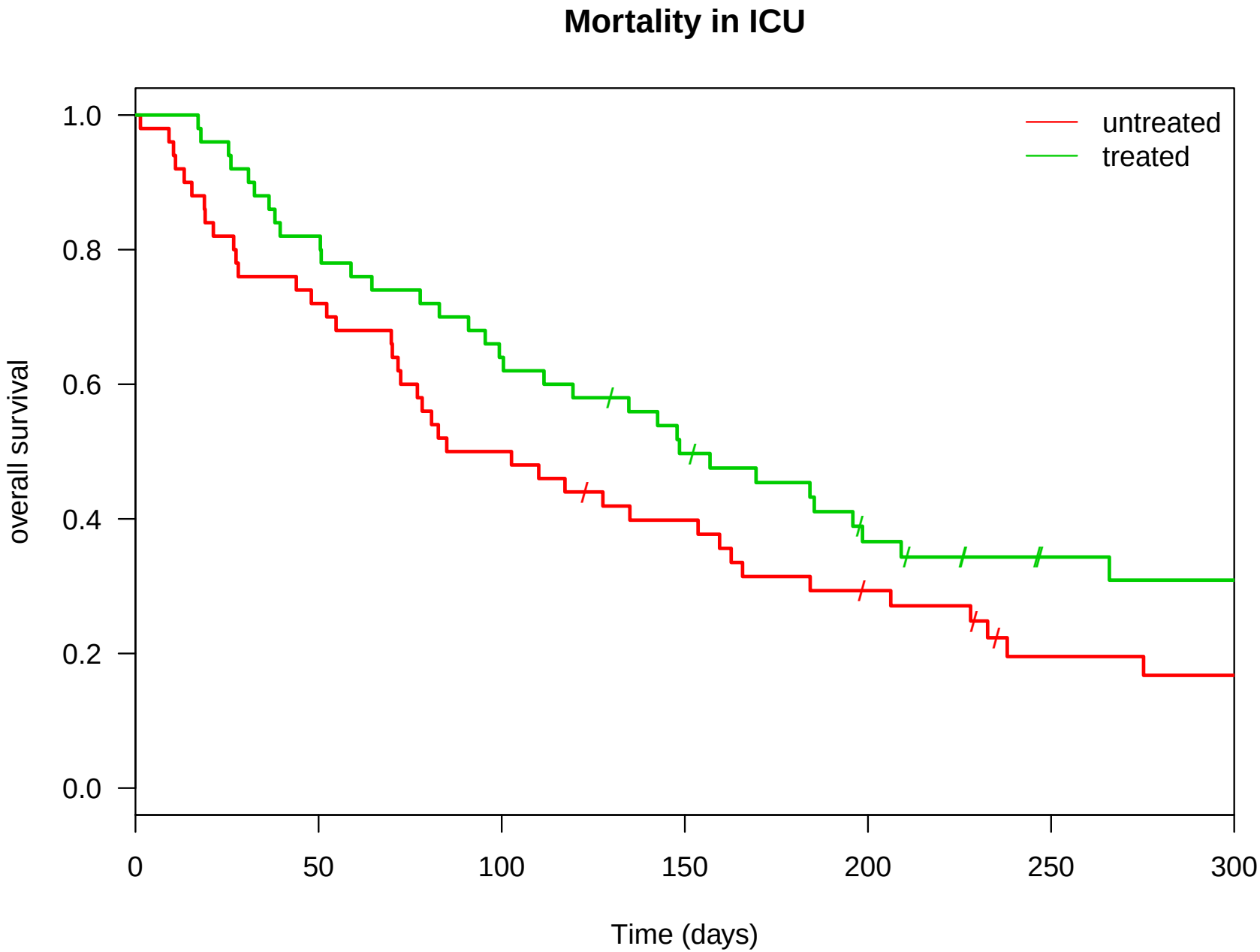
Comparing "no treatment" to "cup of tea" - which would have lower mortality, as assessed e.g. by plotting Kaplan-Meier estimates? **The treatment**!

Why?

- Patients receiving "treatment" cannot die within first 15 days. Kaplan-Meier estimates at 100% until that time. **Immortal bias**.
- Being alive at day 15 is a "marker" of prognostically favourable patients. **Selection bias**.
- Causality: would giving tea at day 15 increase survival for "treated patients"?

No!

Exemplary Kaplan-Meier estimates



Agenda

1. Immortal bias
2. Pharmacometric examples
3. But FDA runs them anyway!
4. What should be the role of exposure - response analyses?
5. Is bias inevitable? And if yes, which direction?
6. Other methods
7. Conclusions

Immortal bias

Immortal bias

Immortal time:

- Period of follow-up during which, by design, the event of interest cannot occur.
- Patients not at risk $\Rightarrow$ **immortal** in that period.
- Any analysis which treats variables assessed **post - baseline** as known at baseline will be subject to **immortal bias**. Bias may be large or small.
- Immortal bias often induces selection bias. Conceptually not easy to keep them apart. For an attempt see e.g. Hernan et al. (2004).

Anderson et al. (1983), Anderson (2001), Anderson et al. (2008).

> *The usual methods of comparing responders to non-responders is* **wrong** *and* **should never be used**.

Immortal bias

Aalen et al. (2015):

> *Paradoxically it is* **well-known** *that in clinical trials one* **should not carry out treatment comparisons by conditioning on variables realized post - randomization** *which may be responsive to treatment since they may be on the causal pathway to the response of interest [...]. Treatment comparisons based on subgroups of individuals defined post - randomization ... are widely known to yield* **invalid** *inferences regarding treatment effects because of the* **benefit of randomization** *is lost in such comparisons.*

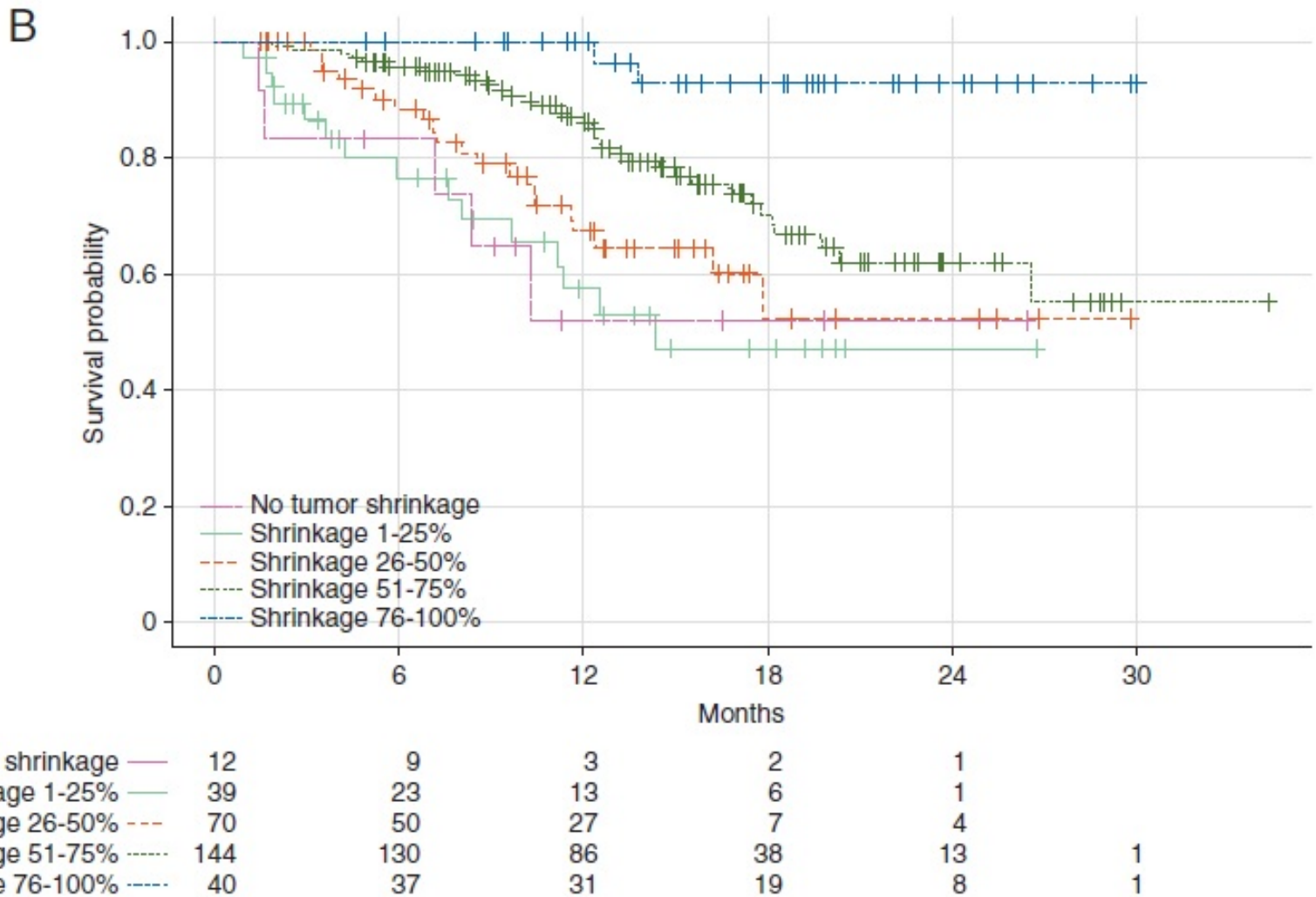
Immortal bias

van Walraven et al. (2004):

- Survival analyses in major medical journals (1998-2002).
- Almost 20% of all analyses contained a time-dependent exposure.
- More than **40%** of these erroneously treated them as known at baseline ⇒ **overestimating** the effect (see below).

Also FDA...

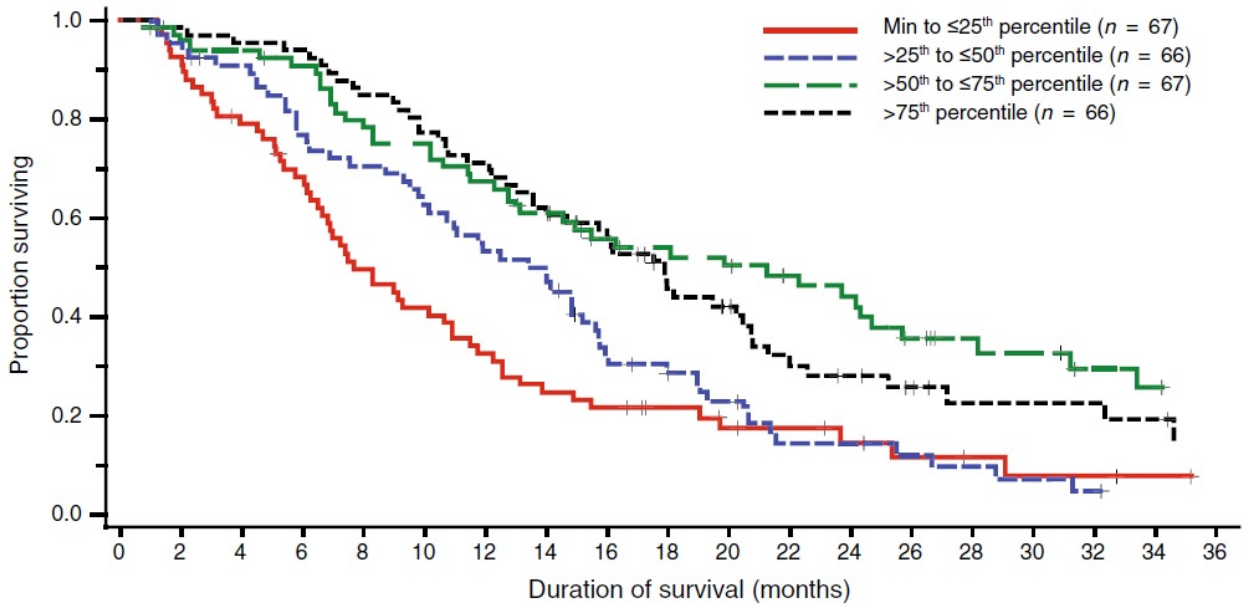
McCoach et al. (2017).



Pharmacometric examples

Example 1: Time-to-event by exposure

Time-to-event by exposure: Herceptin example



Number at risk:

Min to ≤25th percentile	67	62	52	44	32	27	21	16	14	11	8	7	5	4	3	2	2	1	0
>25th to ≤50th percentile	66	63	58	49	45	40	34	31	19	15	12	7	7	5	4	3	2	1	1
>50th to ≤75th percentile	67	62	61	58	50	48	43	38	31	29	27	23	21	16	12	11	8	7	6
>75th percentile	66	65	63	62	56	51	47	40	35	26	22	15	13	10	7	7	7	6	4

ToGA, Cosson et al. (2014):

- PK trial in **Herceptin** program.
- OS by Cycle 1 trough concentration.
- Lowest quartile $\Rightarrow$ shorter survival?
- Baseline characteristics were looked at. Conclusion: *...it is unclear whether the lower OS is due to low drug concentration or to disease burden.*

OS by exposure: Herceptin example

ToGA:

- Concentration measured during Cycle 1, i.e. **post - baseline**. Potential of **immortal bias**.
- **Selection bias**: low average concentration potential marker for patients with unfavourable prognostic profile.
- PK modelling: higher dose $\Rightarrow$ C_{trough} increases in lowest quartile.

How is this to be interpreted?

- **Causally?**
- "Increase mean concentration to increase OS": is that the suggested implication? Unclear, to say the least.
- Suggesting this "implication" might cause trouble: **Post-approval commitment**: HELOISE trial.

HELOISE trial

FDA re-iterated that PAC was justified, Yang et al. (2013):

> *In conclusion, a combined exposure-response and case-control analysis played an important role in identifying a subgroup that may not benefit from trastuzumab under the current regimen. The results of this analysis* **justified** *the FDA recommendation of conducting postmarketing clinical trials to* **investigate a dosing regimen with higher exposure** *[...] and to prospectively evaluate whether this regime will result in acceptable OS benefit.*

Ironically, the goal of their proposed method is...

> *To* **reduce the bias** *introduced by confounding risk factors.*

HELOISE trial

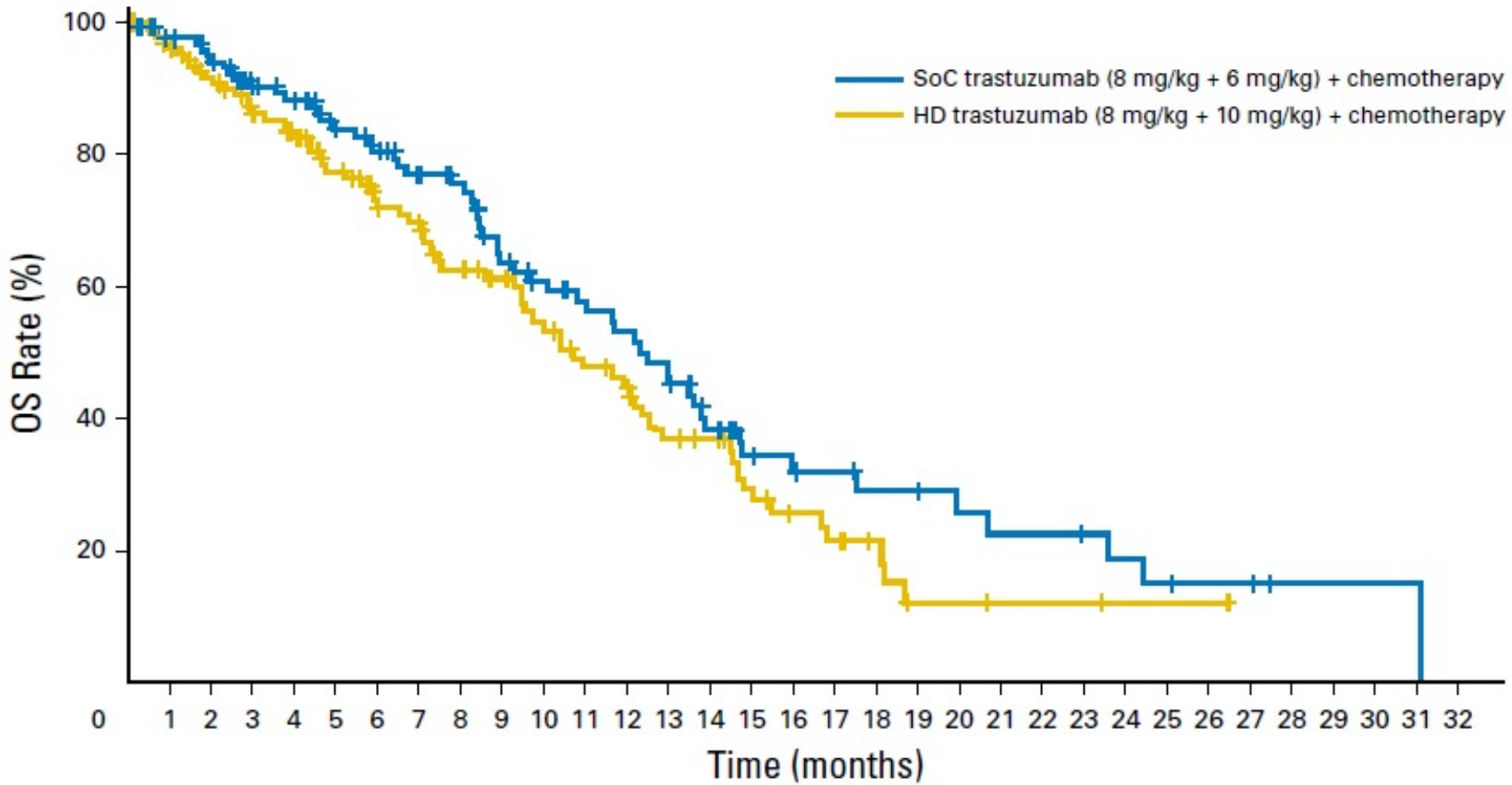
HELOISE, Shah et al. (2017):

> *It has been previously* **demonstrated** *that patients with low cycle 1 trastuzumab $C_{through}$ (eg, fast trastuzumab clearers) had worse overall outcomes.*

Causality implied?

- **RCT** standard of care vs. higher dose.
- 248 patients.
- $C_{through}$ increased.
- Futility interim analysis:

HELOISE trial



HELOISE trial

HELOISE, Shah et al. (2017):

> *It has been previously* **demonstrated** *that patients with low cycle 1 trastuzumab $C_{through}$ (eg, fast trastuzumab clearers) had worse overall outcomes.*

Causality implied?

- **RCT** standard of care vs. higher dose.
- 248 patients.
- $C_{through}$ increased.
- Futility interim analysis: OS hazard ratio **1.24**, 95% CI from 0.86 to 1.78.

> *The apparent exposure OS relationship based on the single dose TOGA trial appears to be a* **confounding effect** *of drug clearance along with poorer clinical factors, rather than a causal exposure-response relationship.*

Kagedal et al. (2017)

Example 2: Time-to-event by tumor growth inhibition metrics

OS by TGI metrics: Avastin example

Han et al. (2016):

- **Tumor growth inhibition** (TGI) metrics: based on *Models for* **longitudinal tumor size** *data.*
- Goal: predict OS with these.
- Post - baseline measurements.

Mistry (2016):

- *...relationship seems* **incredibly strong**, *maybe* **too good to be true**. *Perhaps it could well be [...]. One of the key forms of bias when using covariates that are time-dependent, which TTG and, in fact, any model-derived metrics are, is* **immortal bias**.
- *...Kaplan-Meier curves [...] are* **incredibly misleading and biased**.

Claret et al. (2017):

- *The authors contend that model-derived TGI metrics are not time-dependent and* **not subjected to immortal bias**.

I disagree.

But FDA runs them anyway!

FDA must not be right

McCoach et al. (2017):

- **8/10** authors from FDA, including:
  - Director of the FDA's Oncology Center of Excellence (Pazdur),
  - Director, Division of Biometrics V Office of Biostatistics, Center for Drug Evaluation and Research (Sridhara).
- *Our analysis suggests a greater DepOR [% of maximal tumor reduction from baseline of a target lesion] is associated with longer PFS and OS for patients receiving ALKi or anti-PD1 Ab. Overall, this suggests that DepOR may provide an additional outcome measure for clinical trials, and may allow better comparisons of treatment activity.*

Weber et al. (2018):

The analysis [...] is **prone to immortal bias**, *because DepOR develops over time, but patients were categorized based on DepOR into responder groups using the maximal tumor shrinkage.*

Ignoring time dependency leads to **seriously biased results and therefore to wrong conclusions**.

What should be the role of exposure - response analyses?

Role of exposure - response analysis?

Key question: What scientific question do we want to answer with this analysis?

Exposure - response analysis:

- Typically subject to **immortal** and **selection** bias.
- **Exaggeration** of potential effect.
- Causal conclusions unclear, not to say impossible.
- These limitations should be **clearly stated**!
- May serve as supportive, **descriptive** analysis. Causality statements need to come from alternative analyses.

Role of such analyses, and what conclusions can be drawn from them, needs to be **clarified**.

Is bias inevitable? And if yes, which direction?

Is bias inevitable? And if yes, which direction?

Beyersmann et al. (2008):

> *Biased effect estimation is a* **mathematically inevitable** *consequence of time-dependent [immortal] bias.*
>
> *...there can be no loophole avoiding the estimation bias that follows time-dependent bias.*

Direction of the bias:

1. No effect of the time-dependent exposure on the time until the trial endpoint $\Rightarrow$ biased analysis will show a **prolongation**. ICU & tea example!
2. Prolonging effect $\Rightarrow$ biased analysis will show an **even greater prolongation**. Potentially response - OS, exposure - PFS / OS.
3. Accelerating effect $\Rightarrow$ biased analysis will show at least a less pronounced acceleration.

Explanation: look at hazard estimates in correct and "immortal biased" multistate model. Argument then based on comparing simple proportions.

Other methods

Other methods

1. Only **use baseline covariates**. Often, interest in post - baseline variables as well.
2. **RCT**, e.g. to compare different dosing as in HELOISE. Typically not realistic and/or not desired.
3. **Multistate** models: model transitions between different states explicitly.
    - Connects canonical oncology endpoints response, progression, death.
    - Information before reaching a state can be used as baseline covariate to model transition out of this state $\Rightarrow$ way to incorporate TGI metrics.
    - See Beyer et al. (2018) for an application.

Other methods

4. **Landmark** analyses, e.g. for exposure - PFS:
   - Construction:
     - Set **landmark** at 6 months $\Rightarrow$ new baseline.
     - Only consider patients with $\geq$ 6 months observation time, i.e. remove patients who died / were censored $<$6 months.
     - Mean exposure during months 0 - 6 is then new **baseline** variable.
     - Choice of landmark might be arbitrary.
   - Conditional on landmark status $\Rightarrow$ potentially present initial randomization "lost".
   - Fixes immortal bias, e.g. in **ICU & tea** example.
   - Still subject to selection bias. Adjust using multiple regression or propensity scores.
   - Can answer question of type: *Low exposure in $<$ 6 months is prognostic for time-to-event, with potential adjustment.*

HELOISE: treatment until PD. "Canonical" landmark time less clear.

Other methods

5. Cox regression with **time-dependent covariate**:
   - All patients start as non-exposed.
   - Exposure status may change over time.
   - Use time-dependent covariate in analysis, e.g. in Cox regression model.
   - Advantages over landmark analyses:
     - No choice of **potentially arbitrary landmark**.
     - **Flexible** model of effect of time-dependent exposures on the hazard function.
     - Easy to fit.
   - Ok for hazards. Not (easily) usable for probabilities, i.e. survival functions.
   - Causal interpretation again unclear:
     - Randomization lost.
     - Fixes immortal, but not (necessarily) selection bias.
     - Post - baseline covariates may be on causal pathway between randomized treatment and time-to-event endpoint.

Other methods - advanced

6. **Causal models** to account for (measured) time-dependent confounding:
   - Inverse probability weighted estimation of marginal structural models $\Rightarrow$ Cox regression with time-dependent weights.
   - Valid under **strong assumptions** (no unmeasured confounding).
   - See e.g. Daniel et al. (2013).
7. **Dynamic prediction** including longitudinal covariates: extension of landmarking using **multiple landmarks**. See e.g. van Houwelingen and Putter (2011).
8. **Joint models** of longitudinal covariates and time-to-event e.g. using shared random effects:
   - Model selection bias explicitly.
   - See e.g. Rizopoulos (2012).

Conclusions

Conclusions

- **What scientific question do we want to answer?** Transparently state that!
- Treating post - baseline information as known at baseline: common but subtle.
- **Immortal** and **selection** bias still prevalent in literature, and also FDA analyses.
- Immortal bias inevitably leads to **overestimation of effect** of exposure.
- These biases can be present in **observational studies** and **RCTs**.
- Exposure - response analysis:
  - Make limitations **transparent**.
  - Causal conclusions likely not possible.
  - Typically interpreted as **descriptive**.
  - Risk of **overinterpretation** $\Rightarrow$ post-approval commitment!
- Valid methods to draw causal conclusions exist:
  - Typically make (strong) assumptions.
  - Construction and fitting potentially challenging, e.g. for joint models.

Outlook

We condition on **intercurrent** event in language of ICH E9 estimand addendum.
Framework can help to clarify question one is interested in.

Epidemiological literature offers further alternatives for observational studies, see e.g. Murray and Hernan (2016), Murray and Hernan (2018).

Thank you for your attention.

References I

- Aalen, O. O., Cook, R. J. and Røysland, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Anal* **21** 579–593.
- Anderson, J. (2001). Commonly misused approaches in the analysis of cancer clinical trials. In *Handbook of Statistics in Clinical Oncology* (J. Crowley, ed.), 1st ed. Dekker, New York, 525–542.
- Anderson, J. R., Cain, K. C. and Gelber, R. D. (1983). Analysis of survival by tumor response. *J. Clin. Oncol.* **1** 710–719.
- Anderson, J. R., Cain, K. C. and Gelber, R. D. (2008). Analysis of survival by tumor response and other comparisons of time-to-event by outcome variables. *J. Clin. Oncol.* **26** 3913–3915.
- Beyer, U., Dejardin, D., Meller, M., Rufibach, K. and Burger, H. U. (2018). A multistate model for early decision making in oncology.
- Beyersmann, J., Gastmeier, P., Wolkewitz, M. and Schumacher, M. (2008). An easy mathematical proof showed that time-dependent bias inevitably leads to biased effect estimation. *J Clin Epidemiol* **61** 1216–1221.
- Claret, L., Han, K. and Bruno, R. (2017). Model-Based Estimates of Tumor Growth Inhibition Metrics Are Time-Independent: A Reply to Mistry. *CPT Pharmacometrics Syst Pharmacol* **6** 225.

References II

- Cosson, V. F., Ng, V. W., Lehle, M. and Lum, B. L. (2014). Population pharmacokinetics and exposure-response analyses of trastuzumab in patients with advanced gastric or gastroesophageal junction cancer. *Cancer Chemother. Pharmacol.* **73** 737–747.
- Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G. and Sterne, J. A. (2013). Methods for dealing with time-dependent confounding. *Stat Med* **32** 1584–1618.
- Hernan, M. A., Hernandez-Diaz, S. and Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology* **15** 615–625.
- Han, K., Claret, L., Piao, Y., Hegde, P., Joshi, A., Powell, J. R., Jin, J. and Bruno, R. (2016). Simulations to Predict Clinical Trial Outcome of Bevacizumab Plus Chemotherapy vs. Chemotherapy Alone in Patients With First-Line Gastric Cancer and Elevated Plasma VEGF-A. *CPT Pharmacometrics Syst Pharmacol* **5** 352–358.
- Ho, A. M., Dion, P. W., Ng, C. S. and Karmakar, M. K. (2013). Understanding immortal time bias in observational cohort studies. *Anaesthesia* **68** 126–130.
- ICH E9 working group (2017). ICH E9 (R1): addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. `http://www.ich.org/ichnews/newsroom/read/article/ich-e9r1-revised-guideline-reaches-step-2b-of-the-ich-process.html`

References III

- Kagedal, M., Claret, L., Marchand, M., Chanu, P., Bruno, R., Garg, A. and Jin, J. (2017). Herceptin in her2-positive gastric cancer: Evaluation of exposure-response with two dose levels. Abstract 7329.
  `https://www.page-meeting.org/default.asp?abstract=7329`
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. 2nd ed. John Wiley and Sons, New York-Chichester-Brisbane. Wiley Series in Probability and Mathematical Statistics.
- Marcus, R., Davies, A., Ando, K., Klapper, W., Opat, S., Owen, C., Phillips, E., Sangha, R., Schlag, R., Seymour, J. F., Townsend, W., Trneny, M., Wenger, M., Fingerle-Rowson, G., Rufibach, K., Moore, T., Herold, M. and Hiddemann, W. (2017). Obinutuzumab for the First-Line Treatment of Follicular Lymphoma. *N. Engl. J. Med.* **377** 1331–1344.
- McCoach, C. E., Blumenthal, G. M., Zhang, L., Myers, A., Tang, S., Sridhara, R., Keegan, P., Pazdur, R., Doebele, R. C. and Kazandjian, D. (2017). Exploratory analysis of the association of depth of response and survival in patients with metastatic non-small-cell lung cancer treated with a targeted therapy or immunotherapy. *Ann. Oncol.* **28** 2707–2714.
- Mistry, H. B. (2016). Time-Dependent Bias of Tumor Growth Rate and Time to Tumor Regrowth. *CPT Pharmacometrics Syst Pharmacol* **5** 587.
- Murray, E. J. and Hernan, M. A. (2016). Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials. *Clin Trials* **13** 372–378.

References IV

- Murray, E. J. and Hernan, M. A. (2018). Improved adherence adjustment in the Coronary Drug Project. *Trials* **19** 158.
- Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Biostatistics Series, Chapman & Hall/CRC.
- Shah, M. A., Xu, R. H., Bang, Y. J., Hoff, P. M., Liu, T., Herraez-Baranda, L. A., Xia, F., Garg, A., Shing, M. and Tabernero, J. (2017). HELOISE: Phase IIIb Randomized Multicenter Study Comparing Standard-of-Care and Higher-Dose Trastuzumab Regimens Combined With Chemotherapy as First-Line Therapy in Patients With Human Epidermal Growth Factor Receptor 2-Positive Metastatic Gastric or Gastroesophageal Junction Adenocarcinoma. *J. Clin. Oncol.* **35** 2558–2567.
- van Houwelingen, H. and Putter, H. (2011). *Dynamic Prediction in Clinical Survival Analysis*. Monographs on Statistics & Applied Probability, Chapman & Hall/CRC.
- van Walraven, C., Davis, D., Forster, A. J. and Wells, G. A. (2004). Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol* **57** 672–682.
- Verma, S., Miles, D., Gianni, L., Krop, I. E., Welslau, M., Baselga, J., Pegram, M., Oh, D. Y., Dieras, V., Guardino, E., Fang, L., Lu, M. W., Olsen, S. and Blackwell, K. (2012). Trastuzumab emtansine for HER2-positive advanced breast cancer. *N. Engl. J. Med.* **367** 1783–1791.

References V

- Weber, S., Wolkewitz, M. and Schumacher, M. (2018). Analyzing the impact of depth of response on survival in patients with metastatic non-small-cell lung cancer. *Ann. Oncol.* **29** 282–283.
- Yang, J., Zhao, H., Garnett, C., Rahman, A., Gobburu, J. V., Pierce, W., Schechter, G., Summers, J., Keegan, P., Booth, B. and Wang, Y. (2013). The combination of exposure-response and case-control analyses in regulatory decision making. *J Clin Pharmacol* **53** 160–166.
- Yusuf, S., Wittes, J., Probstfield, J. and Tyroler, H. A. (1991). Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* **266** 93–98.

Backup slides.

Further examples

Time-to-event (T2E) by **chemotherapy dose**:

- "Claim" dose-response effect if **actually administered** high dose associated with T2E.
- Toxicity leading to dose reduction acts as marker of patients with poor prognosis $\Rightarrow$ the longer T2E, the higher the dose.
- Fix: randomize low/high dose.

T2E by **toxicity**: the longer patient's T2E time, the higher odds for tox.

T2E by **compliance to protocol-specified treatment**:

- Compliance may have prognostic importance, irrespective of intervention.
- The longer treatment, the higher odds for non-compliance.

GALLIUM

- **Population**: Treatment-naive follicular lymphoma (FL) patients.
- **Comparison**: Rituximab + chemotherapy vs. **Obinutuzumab** + chemotherapy.
- Rituximab (R): Rituxan, Mabthera. Obinutuzumab: Gazyva(ro) (G).
- **Phase III, 1:1 randomized, open-label** clinical trial.
- 1202 patients.
- Primary endpoint: investigator-assessed **progression-free survival**.
- Treatment paradigm:
  - Chemoimmunotherapy induction for six months.
  - If patient responds: another two years of antibody maintenance therapy.

Marcus et al. (2017), NEJM.

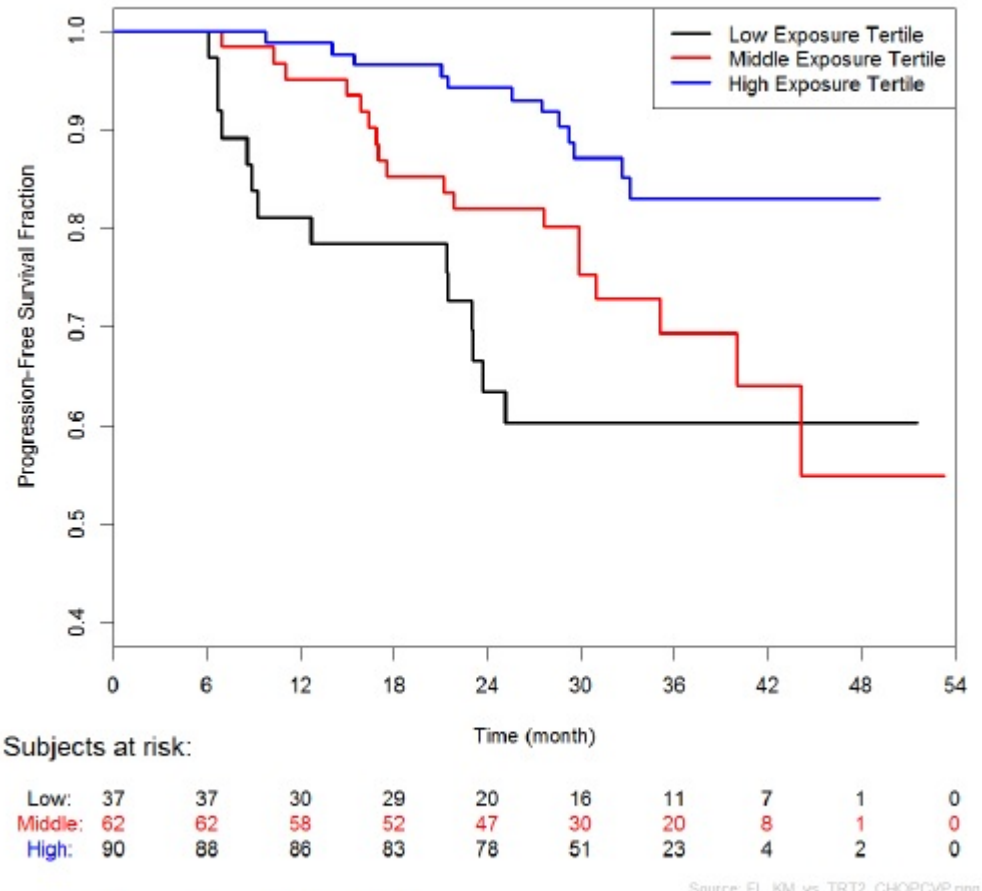
PFS by exposure

Quantification of exposure:

- C_{mean}: mean obinutuzumab concentration over induction period.
- "...patients who received at least half of obinutuzumab induction treatment."
- Variable assessed **post - baseline**.
- Exposure even measured **after** PD for patients progression within first 6 months!
- Groups built by categorizing according to tertiles.

PFS by exposure

Figure 216 Kaplan-Meier Plot of PFS by Tertiles of C_{mean} for Patients with FL on G-CHOP/CVP



Lower tertile: C_{mean} = 68–313 µg/mL;
Middle tertile: C_{mean} = 315–433 µg/mL;
High tertile: C_{mean} = 433–878 µg/mL.

Some patients had event **before** finishing induction treatment.

Be careful

Gallium example:

- Concentration is measured during first six months of treatment, i.e. **post - baseline**. Potential of **immortal bias**.
- **Selection bias**: low average concentration potential marker for patients with unfavourable prognostic profile.

Key question: What scientific question do we want to answer with this analysis?

PFS by exposure

In patients on G-CHOP or G-CVP chemotherapy, the risk of progression or death decreased with increasing exposure. Low exposure (5^{th} percentile of C_{mean}) increased the risk of progression or death by 74% (HR = 1.74), while high exposure (95^{th} percentile of C_{mean}) decreased the risk of progression or death by 61% (HR = 0.394) compared to patients with the median value of C_{mean}.

"...low exposure...**increased** the risk of progression..."

How is this to be interpreted?

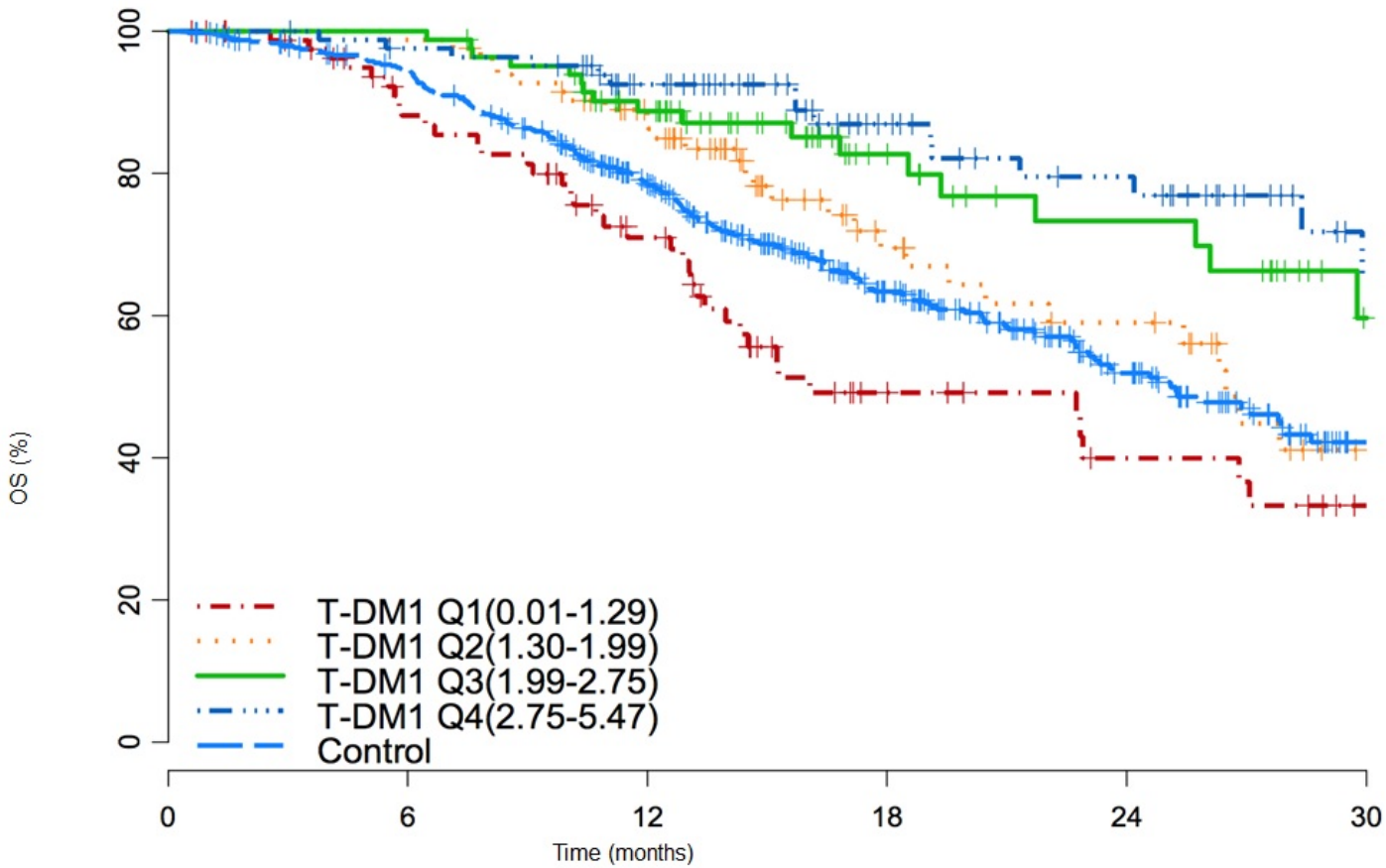
- **Causally?**
- "Increase mean concentration to increase PFS": is that the suggested implication? Unclear, to say the least.
- Suggesting this "implication" might cause trouble $\Rightarrow$ Herceptin experience.

EMILIA trial

Verma et al. (2012).

Exposure-Response (E-R) in Kadcyla / EMILIA:

- FDA conducted E-R analysis on EMILIA efficacy data: *Patients with lower exposure at end of Cycle 1 have lower probability of survival.*
- (Causal!) conclusion rather not justified based on simple analysis.



EMILIA trial

Risk of post - marketing commitment: trial for patients with low exposure?

Team talked FDA out of their "conclusion".

Main argument: Patients with lower exposure **not identifiable at baseline**:

- Neither with baseline covariates,
- nor real-time PK monitoring.

Simple analyses comparing exposure quartiles:

- Subject to immortal and selection bias,
- causal implication unclear,
- Still, Health Authorities draw (causal!) conclusions based on them.

Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 3.5.0 (2018-04-23)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages: survival

This document was generated on 2018-09-04 at 21:15:21.