CONFIDENTIAL

Exposure Metrics Confounding Demo

2023-03-23

Matthew Wiens
Metrum Research Group
Tunxis Road, Suite 112
Tariffville, CT 06081

Contents

1 Introduction

This markdown demonstrates how a biased and incorrect conclusion of the exposure-response relationship can be created by using average concentration through the event (CAVG TTE) instead of more appropriate exposure metrics (e.g. Cycle 1 Average Concentration) which do not depend on the outcome variable.

2 Exposure Simulation

Set up mrgsolve data to generate exposures. The structural model is a 2 compartment model.

We simulate a dose of 100 mg every 3 weeks, for 6 cycles.

```
data <- bbr::nm_join(bbr::read_model(here::here("model/nonmem/106")))

dose_rec <- filter(data, EVID==1)

pars <- distinct(dose_rec,ID,CL,V2,Q,V3,KA,AMT,RF,ACTARM)


dose <- tibble( AMT = 100, ID = 1, TIME=0:5 * 21, EVID=1, CMT=1)   %>%
  inner_join(pars %>% slice_head(n = 1) %>% select(CL, V2, Q, V3, KA),
             by = character()) %>%
  mutate(CL = CL*4)

out <- mrgsim_df(mod,
                 dose,
                 recover = "ACTARM,RF,CL,DOSE",
                 carry.out = "EVID",
                 recsort = 3,
                 tgrid = 0:(6 * 21)) %>%
  as_tibble() %>%
  distinct() %>%
  group_by(TIME) %>%
  filter(EVID != 1) %>%
  ungroup() %>%
```

```
  filter(TIME > 0) %>%
  mutate(CAVG = AUC/TIME)
```

```
typical_concentration_plot <- out  %>%
  ggplot() +
  geom_point(aes(x = TIME, y = CP, color = "Concentration")) +
  geom_line(aes(x = TIME, y = CP, color = "Concentration")) +
  geom_point(aes(x = TIME, y = CAVG, color = "Average Concentration")) +
  geom_line(aes(x = TIME, y = CAVG, color = "Average Concentration")) +
  labs(x = "Time (days)",
       y = "Concentration",
       title = "Typical Concentration Profile",
       color = "Concentration Metric")

typical_concentration_plot
```

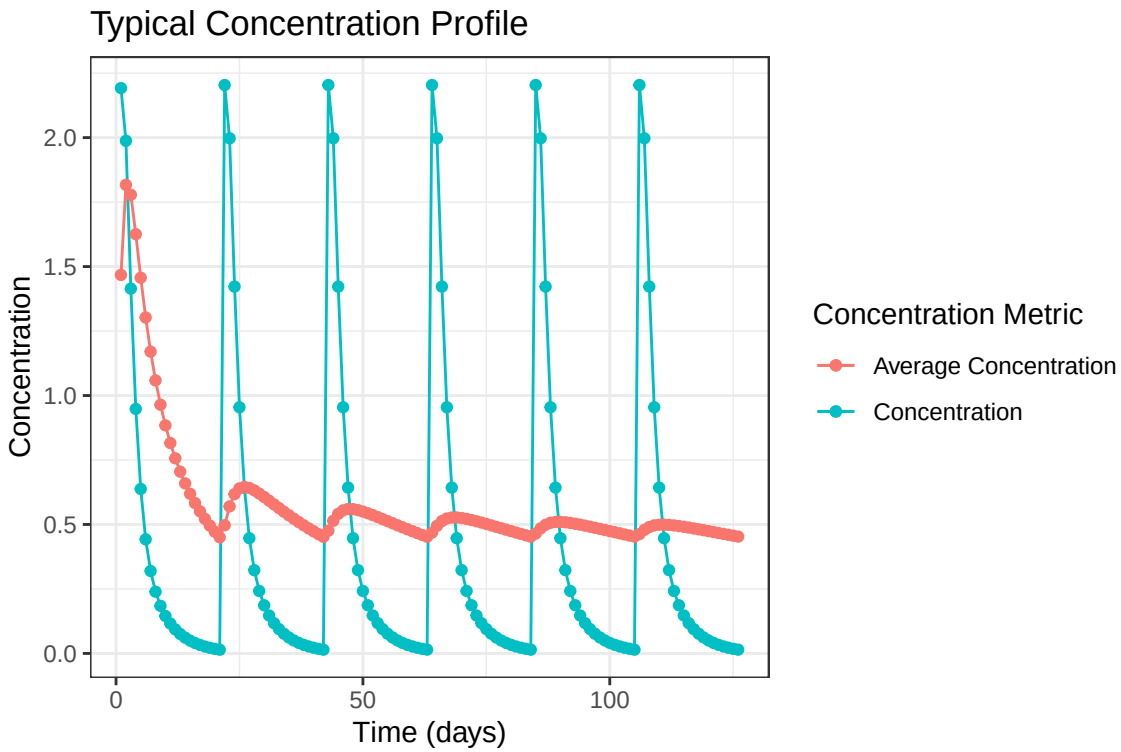

Figure 1: Concentrations for a subject with a reference value of clearance.

```
# Simulate n patients

indiv_dose <- dose %>%
  select(-ID) %>%
  inner_join(tibble(ID = 1:n,
                    ETA1 = rnorm(n, 0, 0.3)),
             by = character()) %>%
  mutate(CL = CL * exp(ETA1)) %>%
  arrange(ID, TIME)
```

```
out2 <- mrgsim_df(mod,
                  indiv_dose,
                  recover = "ACTARM,RF,CL,DOSE",
                  carry.out = "EVID",
                  recsort = 3,
                  tgrid = 0:(6 * 21)) %>%
  as_tibble() %>%
  distinct() %>%
  filter(EVID != 1) %>%
  filter(TIME > 0) %>%
  group_by(ID) %>%
  mutate(CAVG = AUC/TIME) %>%
  ungroup()

CAVGC1_df = out2 %>%
  filter(TIME == 21) %>%
  select(ID, CAVGC1 = CAVG)

exposures = out2 %>%
  inner_join(CAVGC1_df) %>%
  select(ID, TIME, CAVG, CAVGC1) %>%
  arrange(ID, TIME)
```

3 TTE Data

We also simulate TTE data, with **no** dependence on exposure (or anything else). A Weibull survival model is used, visualized below, and we assume 25% of patients won't have the AE regardless of exposure.

```
last_time <- max(exposures$TIME)
```

```
survival_plot <- tibble(x = 0:150) %>%
  mutate(y = pweibull(x, shape = 0.45, scale = 100)) %>%
  ggplot(aes(x = x, y = 0.25 + 0.75 * (1 - y) )) +
  geom_line() +
  labs(x = "Time",
       y = "Survival Probability",
       title = "True Survival Curve") +
  coord_cartesian(ylim = c(0, 1))

survival_plot
```

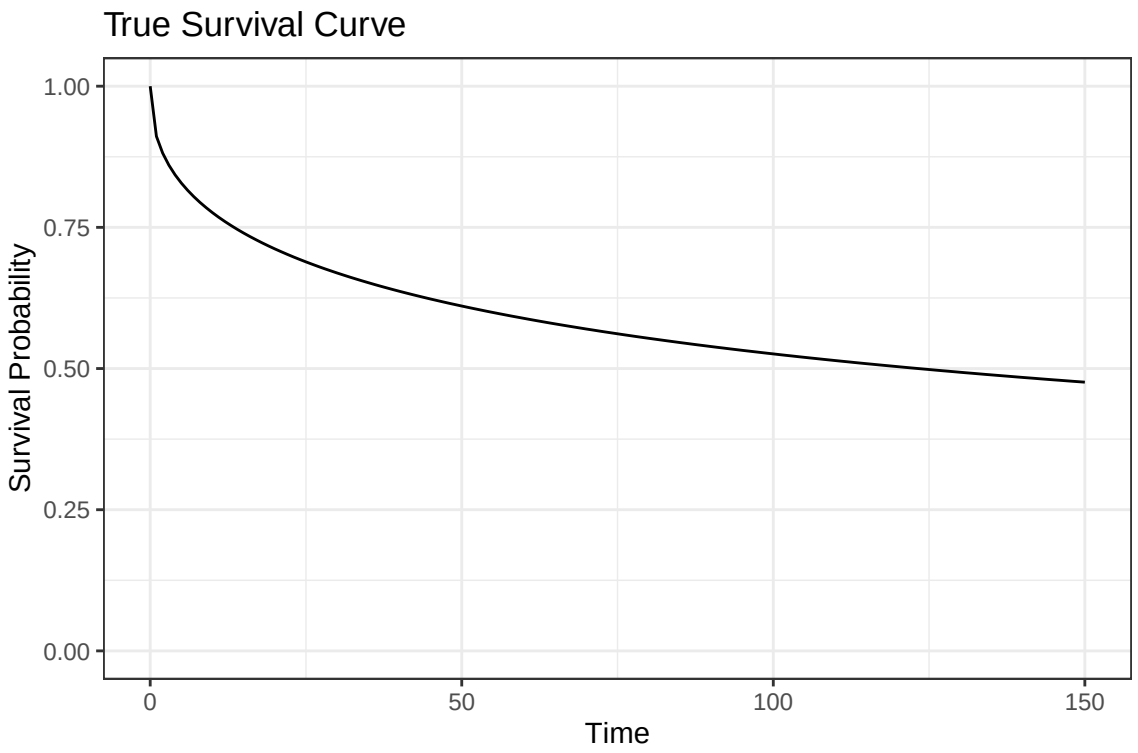

Figure 2: True survival curve used to generate simulated event times

```
TTEs = tibble(ID = 1:n) %>%
  mutate(can_have_event = rbernoulli(n, 0.75)) %>%
  mutate(actual_event_time = 1 + ceiling(rweibull(n, shape = 0.45, scale = 100))) %>%
  mutate(EV = can_have_event * as.numeric(actual_event_time <= last_time) ,
         EVTIME = if_else(can_have_event, pmin(actual_event_time, last_time), last_time))
```

The analysis data is created using simulated exposures and event times. We also derive exposure quartiles for Kaplan-Meier plots.

4 Modeling Results

Three plots of different analysis strategies show the induced confounding between CAVG until the event (CAVG) and probability of having an AE (or survival time until an AE). Logistic regression does ignore potential censoring mechanisms in real problem, but that's not relevant here since the only censoring is adminstrative censoring, at the same day for all subjects.

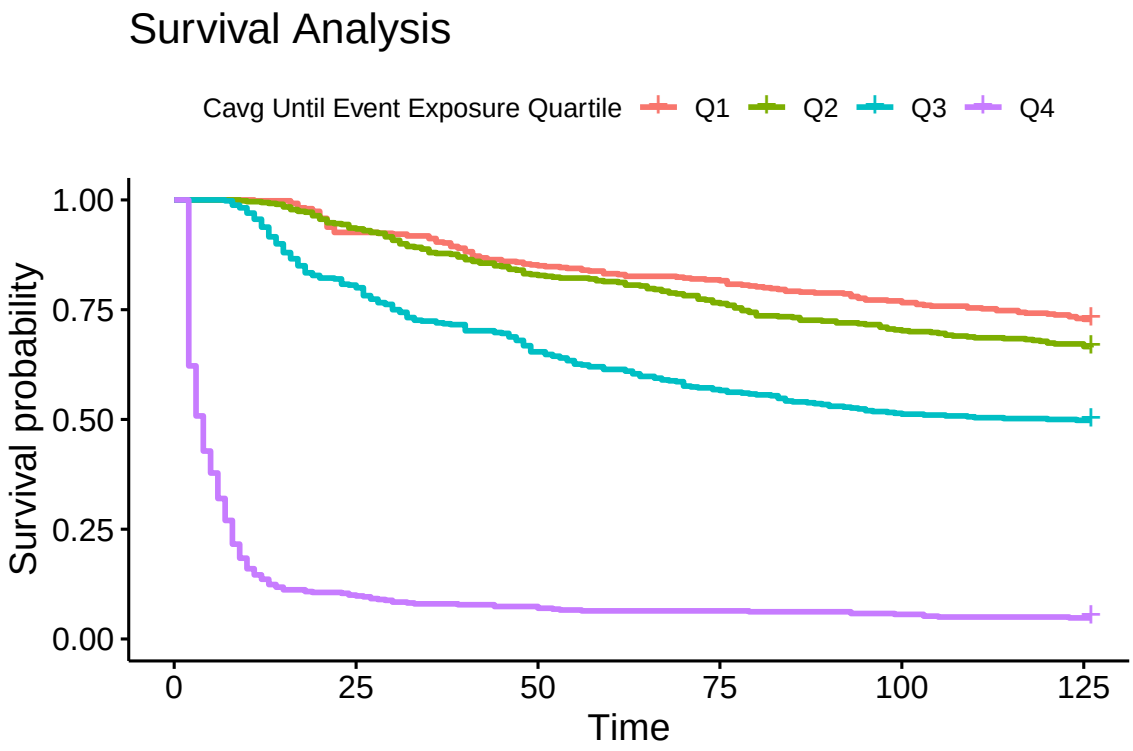

Figure 3: K-M curves by exposure quartile show a clear (biased) E-R relationship

```
cavg_lr <- analysis_dat %>%
  ggplot(aes(x = CAVG, y = EV)) +
  geom_point(shape = "|") +
  geom_smooth(method = "glm", method.args = list(family = "binomial")) +
  scale_x_log10() +
  labs(x = "Cavg Until Event",
       y = "(Probability of) Event",
      title = "Logistic Regression Analysis")
cavg_lr
```

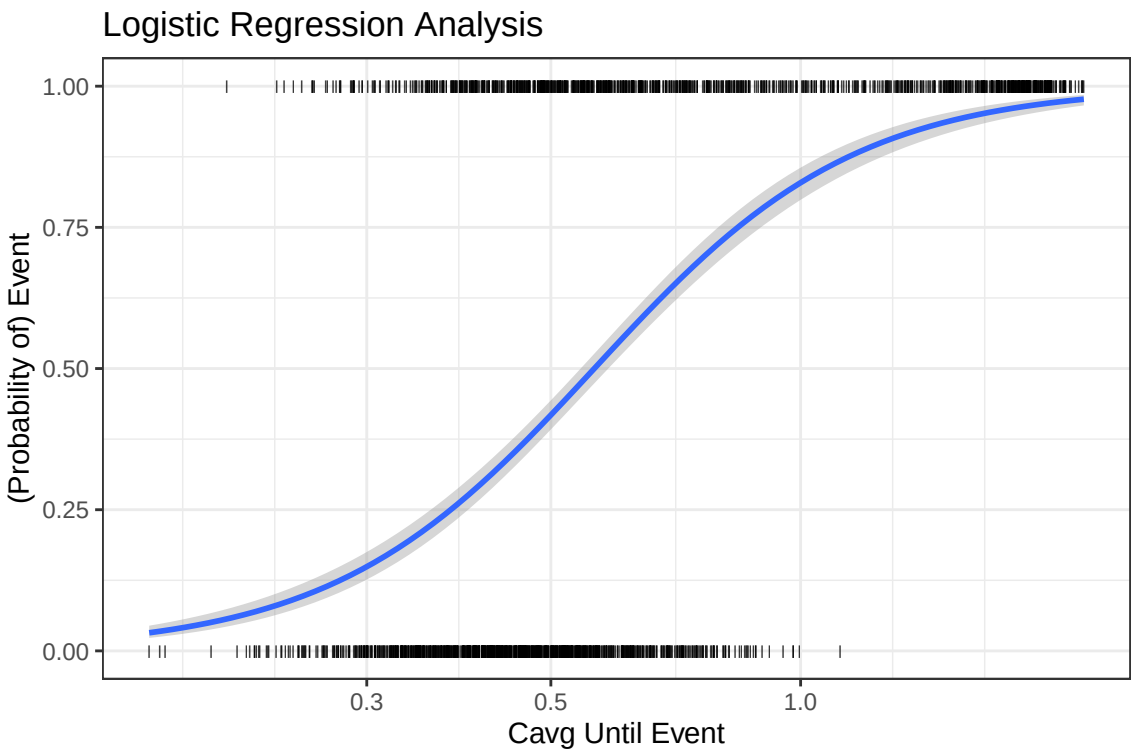

Figure 4: Fitting a logistic regression to the censored simulated data shows a clear (biased) relationship between exposure and risk

Based on this analysis with average concentration, one would conclude that there is an exposure-response relationship, with higher exposures leading to more events. However, we know from the simulation creation that this is **not** the case, and in actuality there is no relationship between exposures and events.

5 No Confouding Analysis

When we use an exposure metric (e.g. predicted cycle 1 average concentration) that does not depend on what we are trying to predict, there is no confounding and we recover the true (null) relationship.

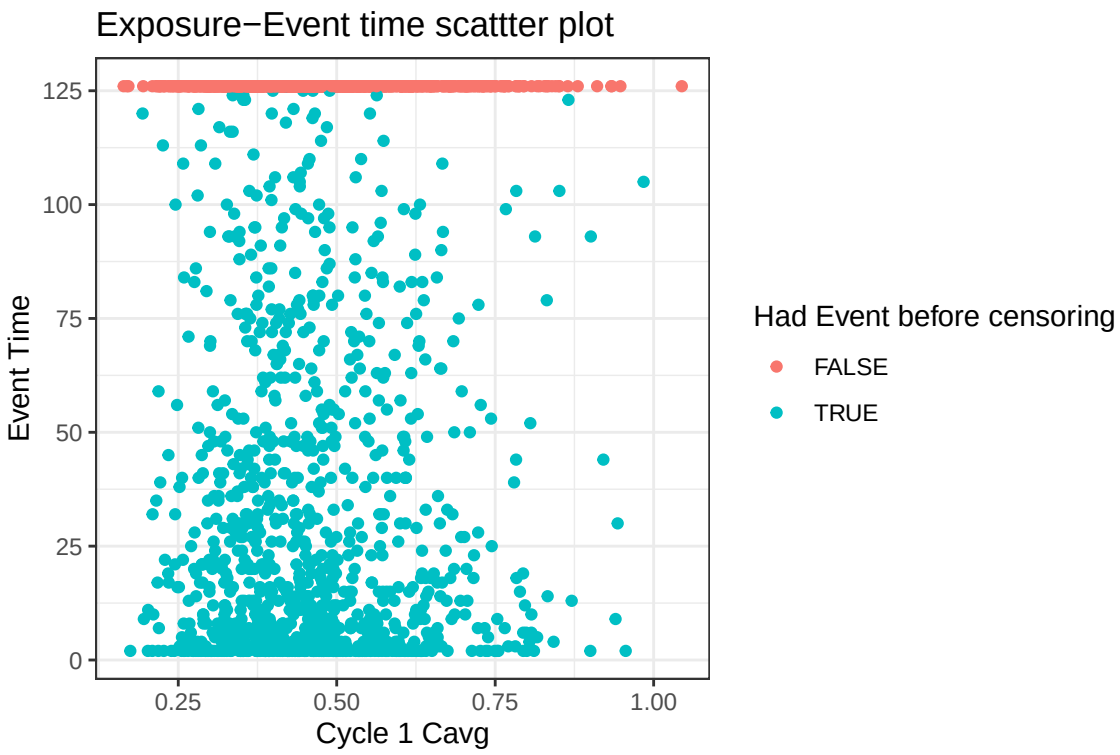

Figure 5: No relationship between Cavg Cycle 1 and event time.

```
c1_km <- ggsurvplot(survfit( Surv(TIME, EV) ~ CAVGC1_quartile, data = analysis_dat),
           legend.labs = c("Q1", "Q2", "Q3", "Q4"),
           legend.title = "Cycle 1 Cavg Exposure Quartile",
           title = "Survival Analysis")

c1_km
```

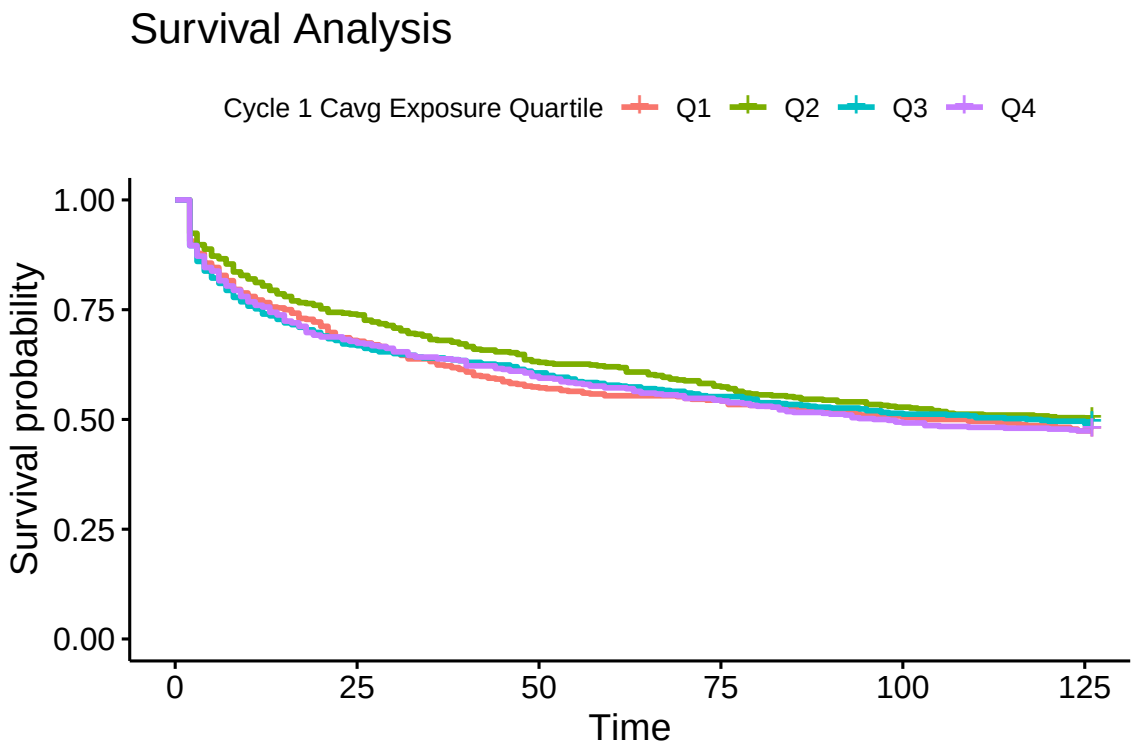

Figure 6: K-M curves by exposure quartile show no clear E-R relationship

```
c1_lr <- analysis_dat %>%
  ggplot(aes(x = CAVGC1, y = EV)) +
  geom_point(shape = "|") +
  geom_smooth(method = "glm", method.args = list(family = "binomial")) +
  scale_x_log10() +
  labs(x = "Cycle 1 Cavg",
       y = "(Probability of) Event",
       title = "Logistic Regression Analysis")

c1_lr
```

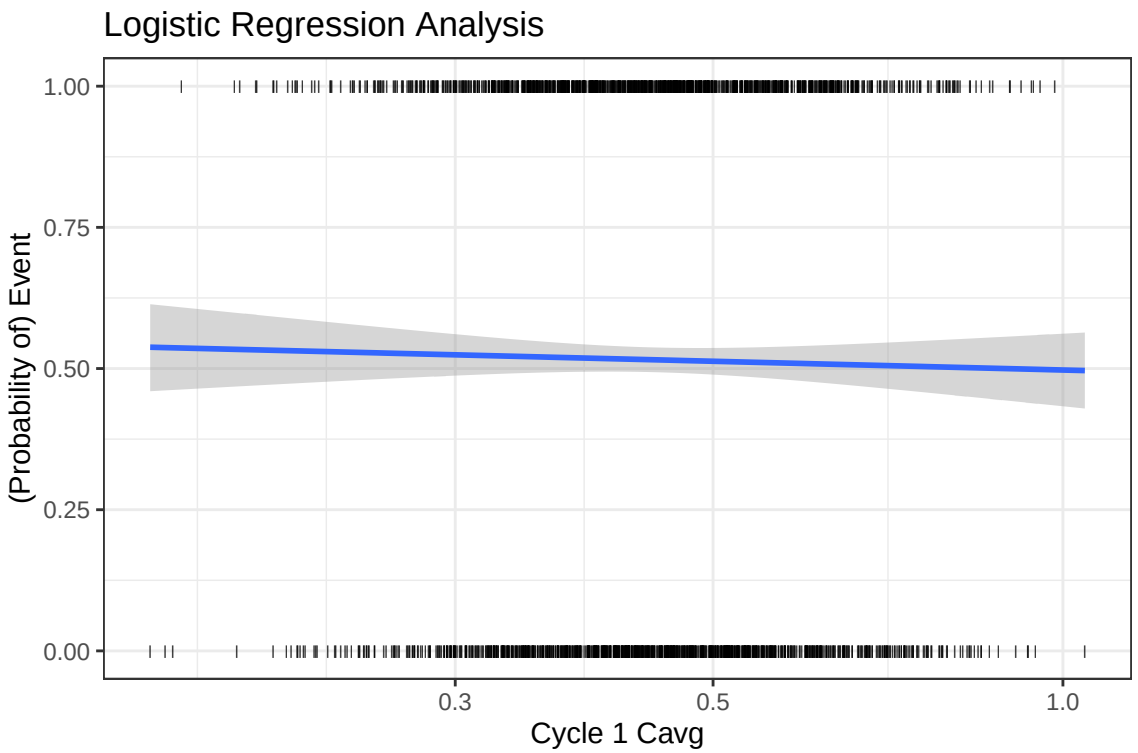

Figure 7: Fitting a logistic regression to the censored simulated data shows no relationship between exposure and risk when using cycle 1 CAVG as the exposure metric

```
## [1] 2046.976
```

```
## [1] 2774.332
```

6 Conclusion

Don't make predictions using predictors that depend on the (time of the) outcome you are predicting, even if the model is "better". Specifically, if two patients who had the exactly the same concentration profile and different event times/status would have different exposure metrics, then there is potential for confounding and biased analysis. Cycle 1 Cavg leads to an unbiased analysis in **this** case, but may not for specific cases (e.g. when clearance depends on tumor size and reductions in tumor size in antibody-drug conjugates)