Assessing the Impact of Measurement Error on B-spline Estimates of Non-Linear Functions: *preliminary results*

Steve Ferreira Guerra & Michal Abrahamowicz
McGill University

(a joint project of STRATOS TG2 and TG4)

STRATOS MEETING

MARCH 29, 2023

Overall Aim

This project is the 1st step in the TG2-TG4 collaboration.

Specific goal :

To assess by simulation the impact of measurement error (ME) in a continuous 'covariate' *X* on its B-spline estimates of its possibly non-linear (NL), relationship *f(X)* with the outcome in logistic regression.

Design of simulation study

- Sample size of N = 1000
- Classical ME model:
  - Observed = Truth + error ($X_i^* = X_i + e_i$)

  - $X \sim Unif(80,150)$

  or

  X resampled from real-world values of SBP at baseline from the Framingham Heart Study

  - $e_i \sim N(0,\sigma_e)$
    - $\sigma_e / \sigma_X = ½$
    - $\sigma_e / \sigma_X = ¾$
    - $\sigma_e / \sigma_X = 1$
- Logistic regression with a Binary outcome : $logit[P(Y=1|X)] = f(X)$
  - Different shapes of true f(X)

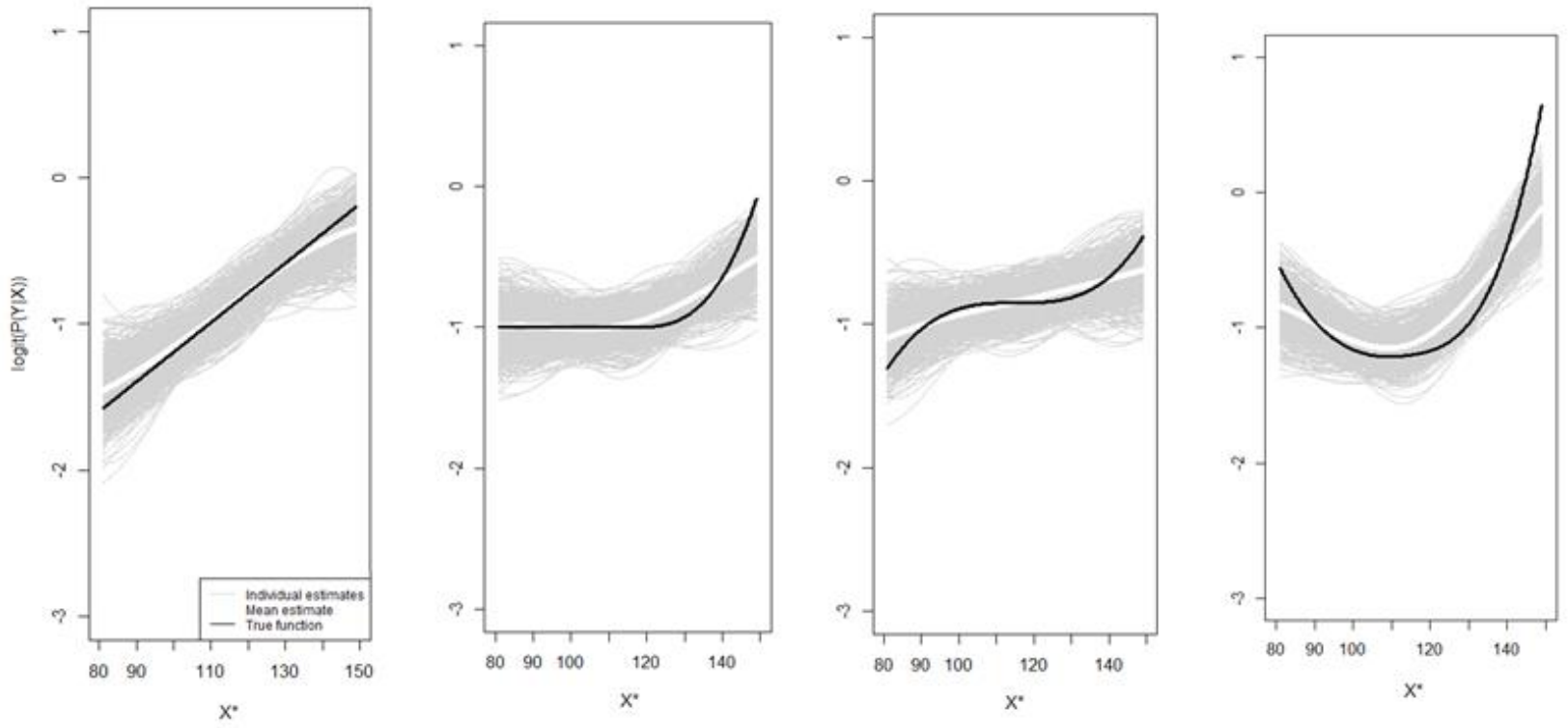
Estimation of f(X*)

Compare estimated curves for:

- Unpenalized cubic regression B-splines with 1 interior knot (4 df) placed at the median of X*

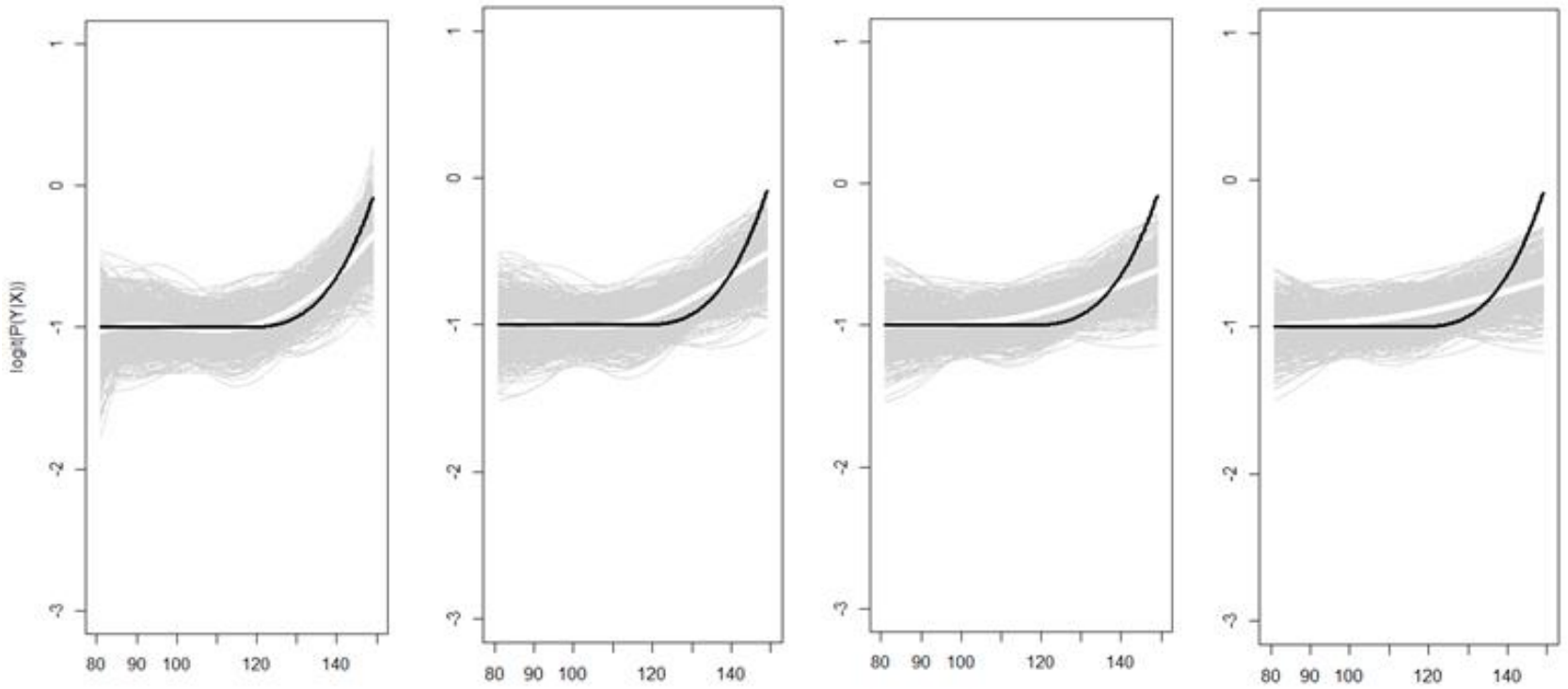
Selected preliminary results

4 different scenarios for true f(X) – (**Black** = True f(X), **Grey** = Individual estimates, **White** = Mean estimate)



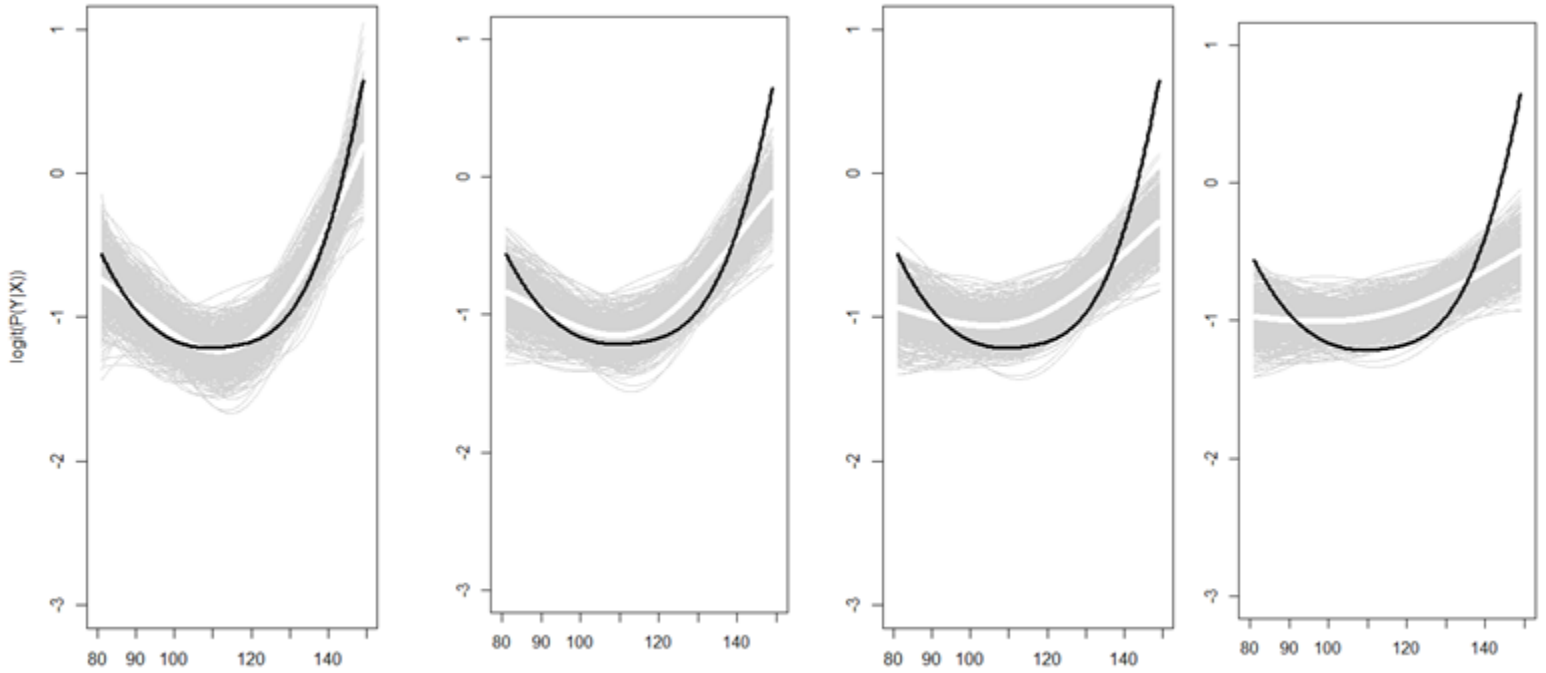
Selected preliminary results

Effect of increasing ME (σ_e = 5, 10, 15, 20)



Selected preliminary results

Effect of increasing ME (σ_e = 5, 10, 15, 20)



Conclusion

Main results (N = 1000 ~ 300 events):

In presence of ME, usually a flattening of the estimated relationship was observed.

→However, over some portions of the X-axis the estimated slope may be higher than the true slope.

→ Increasing ME results in increased linearization with the underlying NL relationship almost completely lost with very large ME.