

# Optimising covariate allocation at design stage using Fisher Information Matrix for Non-Linear Mixed Effects Models in pharmacometrics



Lucie Fayette<sup>@ 1 2</sup> Karl Brendel<sup>2</sup> France Mentré<sup>1</sup>  
@lucie.fayette@inserm.fr<sup>1</sup> Université Paris Cité, Inserm, IAME, F-75018 Paris, France  
<sup>2</sup> Pharmacometrics, Ipsen Innovation, Les Ulis, France



## CONTEXT

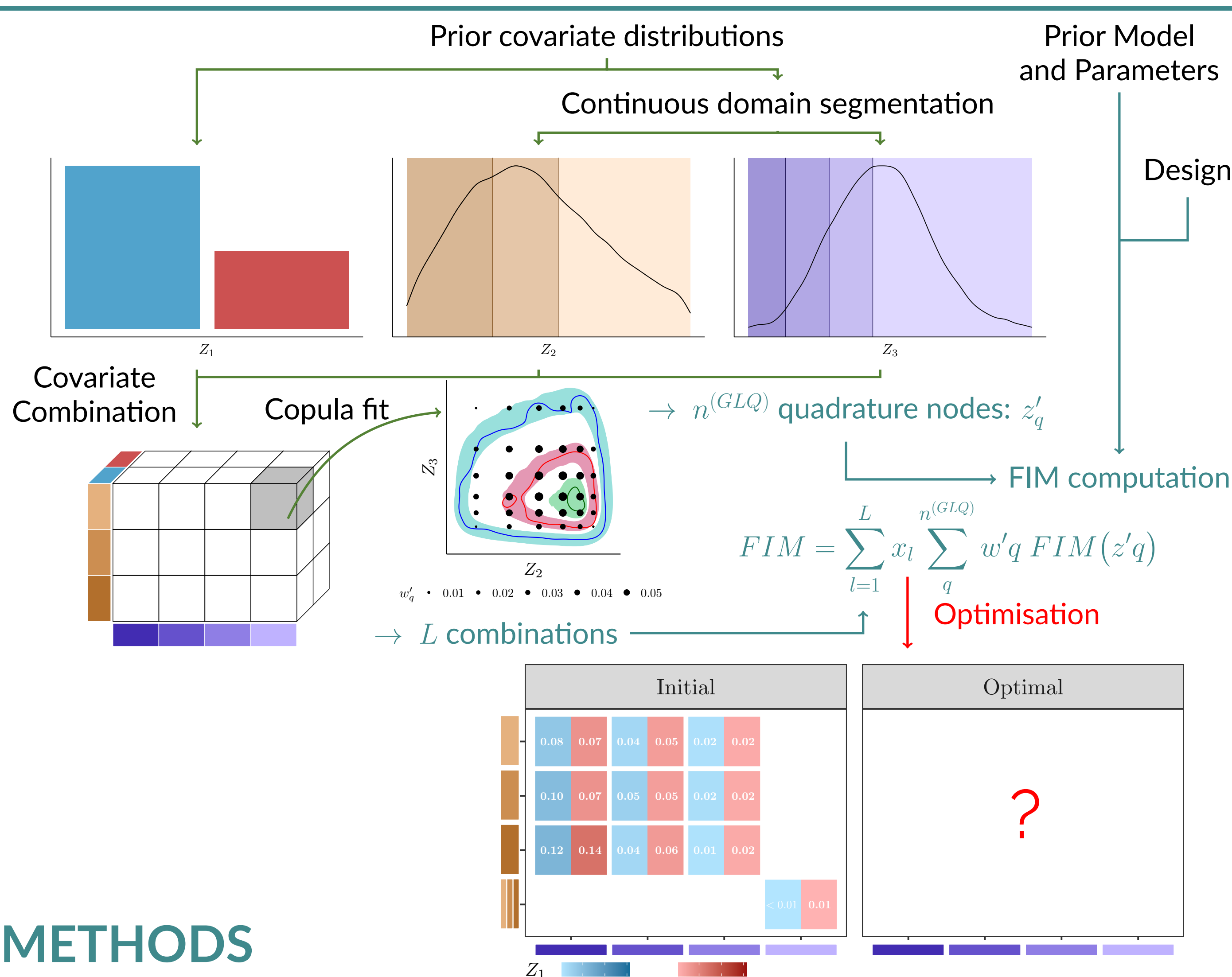
- Aim of pharmacometrics analysis: to identify and quantify covariates that explain part of inter-individual variability (IIV)
- Design optimization: find the design that provides the most informative data for estimating model parameters while respecting some constraints
  - Clinical trial simulations (CTS) for power assessment + sample size computation [1, 2]  
→ Computationally expensive + non-exhaustive
  - Fisher Information Matrix (FIM) → linearization structural model + Gaussian approximation [3–5]
  - Software: ■ PFIM [6] + R package PFIM 6.1 [7] ■ PopED [8, 9] ■ NONMEM\$DESIGN [10]

→ Usually, the distribution of covariates is not considered at this stage

**Goal:** Optimise covariate distribution among subjects to be included in a new study

- Achieve better overall parameter estimations
- Increase the power of statistical tests on covariate effects

## GRAPHICAL ABSTRACT



## METHODS

### FIM computation with covariates [11]

- Given ○ an elementary design  $\xi_i$  ○ a population parameter vector  $\Psi$  ○ a covariate realisation  $z_i$   
 $M_F(\xi_i, \Psi, z_i) = \mathbb{E} \left( \frac{\partial \log l(y_i; \xi_i, \Psi, z_i)}{\partial \Psi} \frac{\partial \log l(y_i; \xi_i, \Psi, z_i)}{\partial \Psi^T} \right)$  where  $l(y_i; \xi_i, \Psi, z_i) = \int p(y_i | \eta_i; \Psi, z_i, \xi_i) p(\eta_i; \Psi, z_i) d\eta_i$
- Given ○ a population design  $\Xi = \{N, (\xi_1, \dots, \xi_N)\}$  with  $N$  subjects ○ a distribution  $p_Z$  for covariates  
 $M_F(\Xi, \Psi, p_Z) = \sum_{i=1}^N M_F(\xi_i, \Psi, p_Z)$  where  $M_F(\xi_i, \Psi, p_Z) = \int M_F(\xi_i, \Psi, z) p_Z(z) dz$

### Covariate distribution optimisation

- Discrete covariates → Optimise the proportion of each category
- Continuous covariates → Segment their domains into intervals, chosen for their clinical meaning → Optimise proportion of each interval in the trial population
  - Preserves the continuous nature of covariates in the FIM computation
  - Discrete segmentation → Reduces dimensionality + Facilitates subject inclusion in clinical setting
- Combining categories × intervals → **L possible distributions for covariate vector Z**

$$M_F(\xi_i, \Psi, p_Z(x)) = \sum_{l=1}^L x_l M_F(\xi_i, \Psi, p_{Z,x_l})$$

where  $x_l$  the proportion of subjects within the trial population having the covariate distribution  $p_{Z,x_l}$

### Computing $M_F(\xi_i, \Psi, p_{Z,x_l})$ : Copula Modelling and Gauss-Legendre Quadrature

- Distribution  $p_{Z,x_l}$  → Continuous covariate modelling with copula
  - Rosenblatt transform [12]:  $T_{Cont, |z_{Dis}}(Z_{Cont}) \sim \mathcal{U}([0; 1]^C)$

$$M_F(\xi_i, \Psi, p_{Z,x_l}) = \int_{[0;1]^C} M_F(\xi_i, \Psi, z_{Dis}, T_{Cont, |z_{Dis}}^{-1}(u_1, \dots, u_C)) du_1 \dots du_C$$

- Gauss-Legendre Quadrature (GLQ) [13]

### Optimisation procedure

- D-criterion  $\phi_D(\Xi, \Psi, p_Z(x)) = (\det(M_F(\Xi, \Psi, p_Z(x))))^{1/P}$  with  $P$  the size of  $\Psi$
- Projected Gradient Descent Algorithm (PGD)
- Number of Subjects needed (NSN) computation

$$NSN(\Xi, \Psi, p_Z(x))_{\beta} = N \left( \frac{SE(\Xi, \Psi, p_Z(x))_{\beta}}{SE^*} \right)^2$$

with  $SE(\Xi, \Psi, p_Z(x))_{\beta}$  standard error on  $\beta$ ,  $SE^*$  level to reach

$NSN_S$ ,  $NSN_R$  and  $NSN_{NR}$  to achieve 80% power for significance, relevance and non-relevance tests

## REFERENCES

1. FDA Guidance for Industry Population Pharmacokinetics. 2022. 2. Sanghavi et al. CPT Pharmacometrics Syst Pharmacol (2024). 3. Mentré et al. Biometrika (1997). 4. Retout et al. Stat Med (2002). 5. Bazzoli et al. Stat Med (2009). 6. Bazzoli et al. Comput Methods Programs Biomed (2010). 7. Leroux et al. PAGE (2023). 8. Foracchia et al. Comput Methods Programs Biomed (2004). 9. Nyberg et al. Comput Methods Programs Biomed (2012). 10. Bauer et al. CPT Pharmacometrics Syst Pharmacol (2021). 11. Fayette et al. medRxiv (2024). 12. Rosenblatt. Ann Stat (1952). 13. Golub et al. Math Comput (1969). 14. CDC.NCHS. National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S.. 2009-2020.

## APPLICATION

### Example: Set-Up for a new trial

- PK Model ○ 1 compartment - IV bolus - linear elimination ○ Combined error:  $y = f + (a + bf)$
- Covariate effects:  $\begin{cases} \ln V_i = \ln \mu_V + \beta_{V,BMI} \ln \left( \frac{BMI_i}{BMI_{pop}} \right) + \beta_{V,SEX} SEX_i + \eta_{V,i} \\ \ln CL_i = \ln \mu_{CL} + \beta_{CL,CLCR} \ln \left( \frac{CLCR_i}{CLCR_{pop}} \right) + \eta_{CL,i} \end{cases}$

Table 1. Population PK parameter values

| Parameter | $\mu_{CL}$ (L/h) | $\mu_V$ (L) | $\omega_{CL}$ | $\omega_V$ | $a$ (mg/L) | $b$  | $\beta_{V,SEX}$ | $\beta_{V,BMI}$ | $\beta_{CL,CLCR}$ | $\tau_{V,SEX}$ | $\tau_{V,BMI}$ (Obesity) | $\tau_{CL,CLCR}$ (Severe) |
|-----------|------------------|-------------|---------------|------------|------------|------|-----------------|-----------------|-------------------|----------------|--------------------------|---------------------------|
| Value     | 0.43             | 7.13        | 0.30          | 0.24       | 2.39       | 0.08 | -0.35           | 0.50            | 0.27              | 0.70           | 1.14                     | 0.70                      |

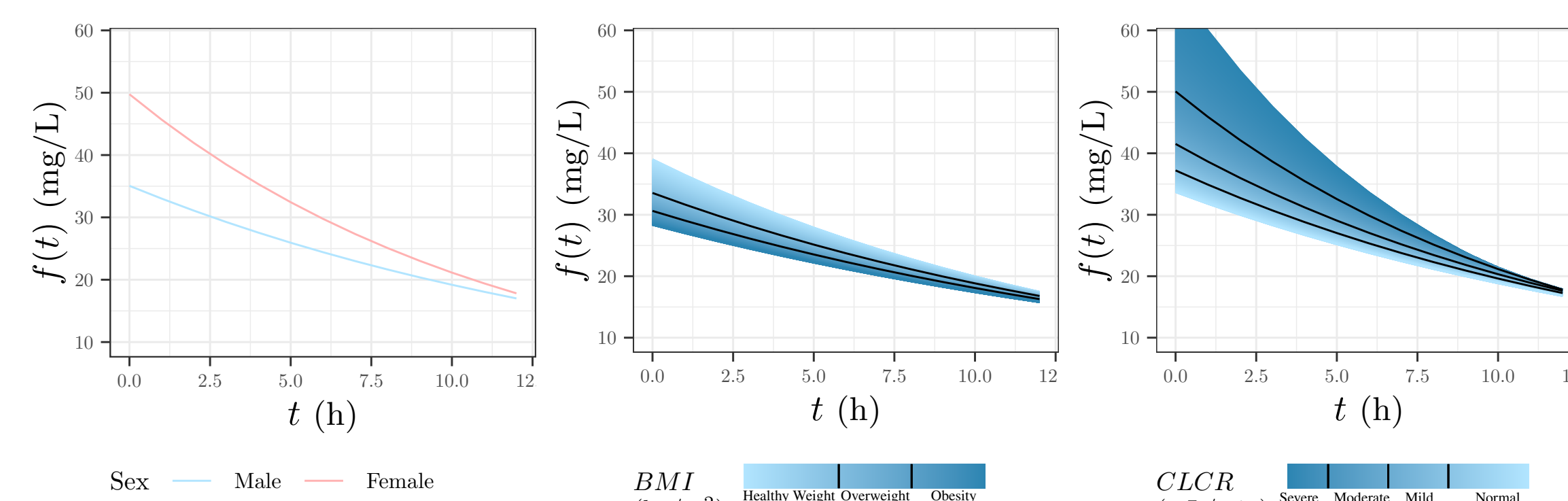


Figure 1. Concentration evolution according to fixed effects, considering only *SEX* (left), *BMI* (middle) or *CLCR* (right)

- Reference area: [0.80; 1.25] → *SEX* and *CLCR*: relevance? - *BMI*: non-relevance?
- Design ○ 250 mg of drug at time 0 ○ 3 samples: 1, 4 and 12 hours post-dose
- National Health and Nutrition Examination Surveys (NHANES) [14] - 2009 to 2020
  - Age between 18 and 80 years old
  - Body Mass Index ( $BMI \geq 18.5$  kg/m<sup>2</sup>)
  - No missing data

→ 29 463 covariate vectors

Table 2. Covariate data summary from the NHANES database, 2009-2020

| Continuous covariates           |             |                   | Continuous segmentation                |                |      |
|---------------------------------|-------------|-------------------|----------------------------------------|----------------|------|
| Covariate                       | mean (sd)   | median [min; max] | Condition on covariate                 | Interval Name  | %    |
| <i>BMI</i> (kg/m <sup>2</sup> ) | 29.5 (7.0)  | 28.2 [18.5; 92.3] | $18.5 \leq BMI < 25$ kg/m <sup>2</sup> | Healthy Weight | 27.9 |
| <i>CLCR</i> (ml/min)            | 99.0 (35.6) | 97.3 [4.1; 342.9] | $25 \leq BMI < 30$ kg/m <sup>2</sup>   | Overweight     | 32.7 |
|                                 |             |                   | $BMI \geq 30$ kg/m <sup>2</sup>        | Obesity        | 39.4 |
|                                 |             |                   | $CLCR \geq 90$ mL/min                  | Normal RF      | 58.4 |
|                                 |             |                   | $60 \leq CLCR < 90$ mL/min             | Mild RF        | 28.3 |
|                                 |             |                   | $30 \leq CLCR < 60$ mL/min             | Moderate RF    | 12.0 |
|                                 |             |                   | $CLCR < 30$ mL/min                     | Severe RF      | 1.2  |

- Combination *SEX* × *BMI* × *CLCR* → 24 combinations
- Because poor representation of Severe RF → Pool in just 2 combinations with *SEX*  
→  $L = 20$  combinations → 20 FIM to compute → 20 proportions to be optimized
- 3 constraints settings:
  - No Constraint: Without any other constraints than proportion definition
  - Lower Constraints: each *BMI* and *CLCR* interval should represent at least 5% of the population
  - Lower and Upper constraints: Lower Constraints + Severe RF should not represent more than 10%

### Optimisation results

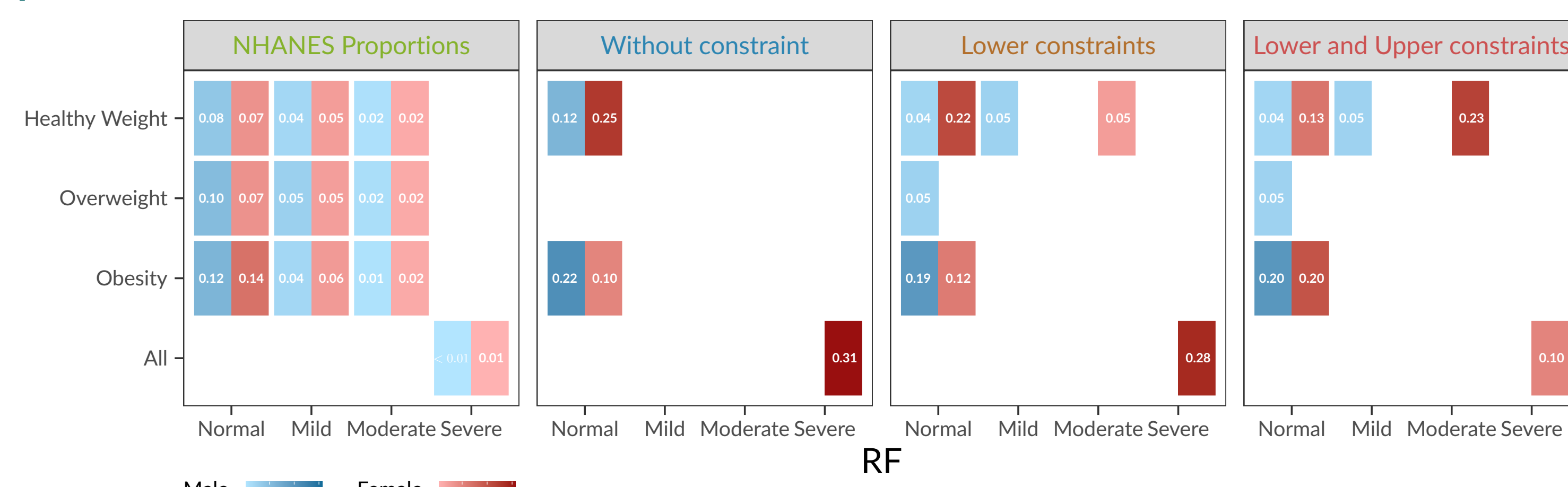


Figure 2. Optimal proportions for the covariate distributions for the D-criterion for the different sets of constraints

- Many categories are not present - Distribution between extreme values
- Female more associated with Healthy Weight while Male associated with Obesity
- Only Female with Severe RF
- Lower Constraints Overweight and Mild + Moderate RF at the 5% minimum
- Lower and Upper constraints Severe RF at the 10% maximum, replaced by Moderate RF

Table 3.  $\phi_D$  and NSN to reach 80% power in covariate testing with the optimal covariate distributions

| Name                          | Optimal designs    |                    |                   |                             |
|-------------------------------|--------------------|--------------------|-------------------|-----------------------------|
|                               | NHANES Proportions | Without Constraint | Lower Constraints | Lower and Upper constraints |
| $\phi_D$                      | 62.47              | 77.1               | 76.11             | 74.56                       |
| $NSN_S(\beta_{V,SEX})$        | 28                 | 33                 | 33                | 33                          |
| $NSN_S(\beta_{V,BMI})$        | 69                 | 57                 | 58                | 52                          |
| $NSN_S(\beta_{CL,CLCR})$      | 106                | 26                 | 28                | 41                          |
| $NSN_R(V, SEX)$               | 167                | 197                | 197               | 196                         |
| $NSN_{NR}(V, BMI$ (Obesity))  | 129                | 106                | 108               | 97                          |
| $NSN_R(CL, CLCR$ (Severe RF)) | 596                | 142                | 153               | 229                         |

→ Even the most constrained optimisation reduces the *NSN* by more than 60%

## Conclusions

- Optimisation method for covariate distribution leading to a reduction in *NSN*
  - Segmentation of continuous domains based on a clinically meaningful definition
    - Continuous range for inclusion
    - Discrete range for inclusion
  - Fast and deterministic computation of the FIM
- Limited by dimensionality: number of covariates + number of nodes  
→ Adaptive methods to reduce the number of nodes?
- Model + parameters + covariate distribution assumed to be known  
→ Extend to a robust approach
- Non-linear constraints? (e.g. test power, *RSE*) → Penalty term, other algorithms?
- Other optimisation criteria (e.g.  $\phi_{DS}$  → Focus on some parameters)
- Joint optimisation of covariates + elementary designs
- Difficult in practice to implement the optimal design