

# Effect of censoring on parameter estimation and predictive ability using an indirect genetic model

## A simulation study

Marzieh Heidaritabar, Ole F. Christensen, Hanne M. Nielsen, Birgitte Ask

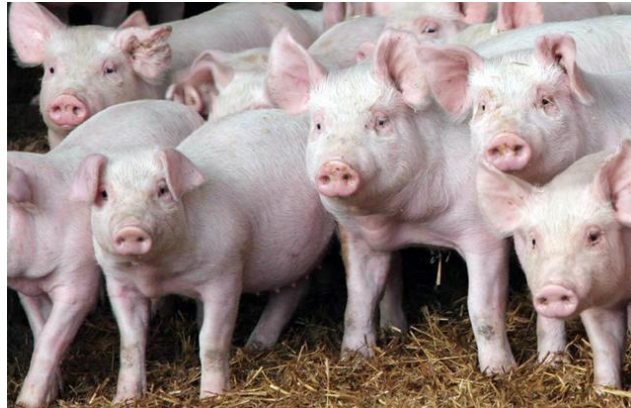
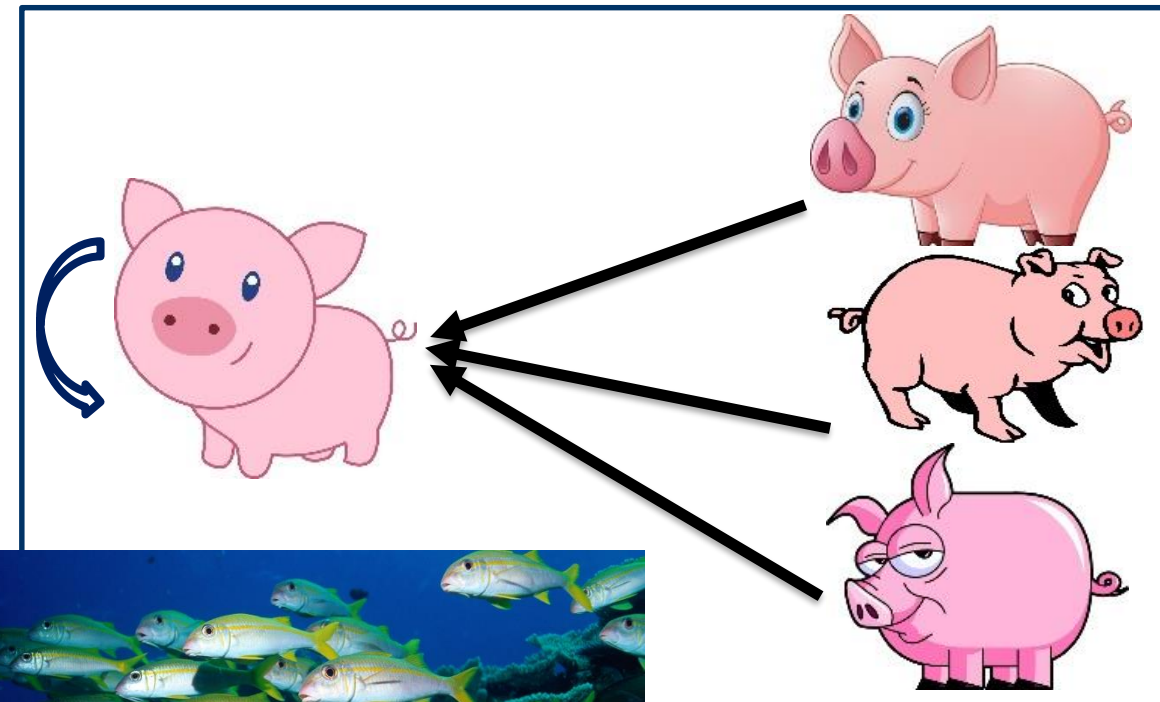
*Presenter: Birgitte Ask*



# Social interactions

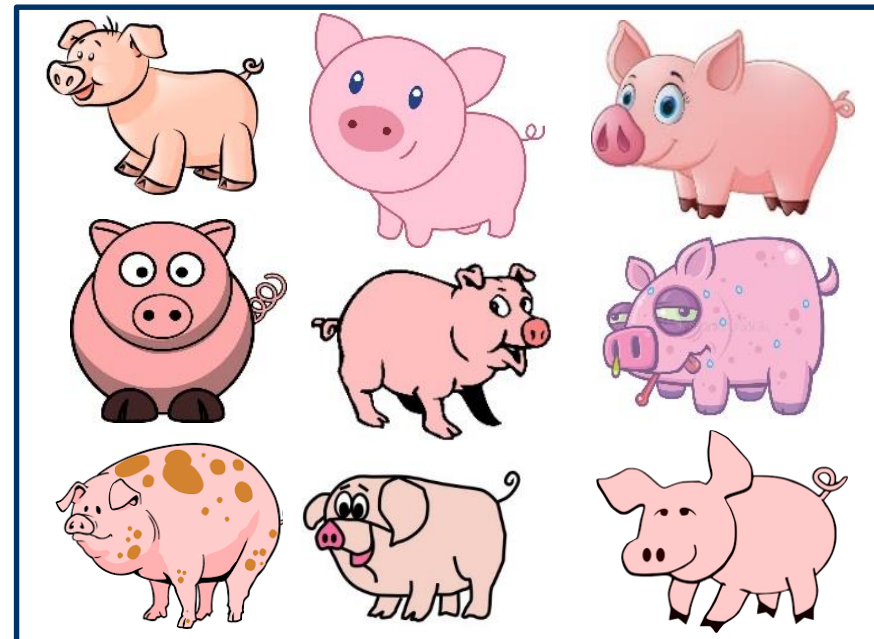
- **Phenotype** = Direct genetic effect (**DGEs**) + Indirect genetic effects (**IGEs**)

- IGEs are important in group-housed animals



# Censoring

- Incomplete observations for traits (censoring)
  - Mortality from diseases
  - Involuntary culling



# Censoring and IGM

- Censoring problematic for indirect genetic models (IGM)?
  - Change in group size with censoring
  - Impact on phenotypes
  - Non-random removals (related to IGEs)
- Possible effect on:
  - estimation of the indirect genetic variance?
  - predicted genetic effects from IGM?

# Objective

**To evaluate the effect of censoring on:**

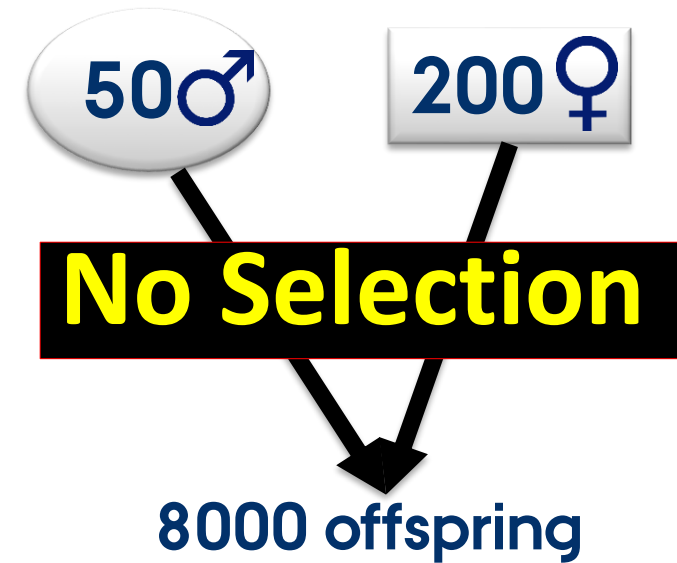
- 1. The estimated indirect genetic variance**
- 2. The predictive ability of an IGM**

# Simulated Scenarios

Data	Description
1. $\text{dat}_{\text{full}}$	Complete data
2. $\text{dat}_{\text{train}}$	<b>50% groups</b> in <b>training</b> had a missing value
3. $\text{dat}_{\text{valid}}$	<b>50% groups</b> in <b>validation</b> had a missing value
4. $\text{dat}_{\text{train\_valid}}$	<b>50% groups</b> in <b>both training and validation</b> had a missing value

# Simulated Data

- **Pedigree:** 6 generations (48000 individuals)
- **Group assignment:** Random
- **Group size:** 8
- **Phenotypes** in the last 2 generations:
  - 5<sup>th</sup> generation: training data (1000 groups)
  - 6<sup>th</sup> generation: validation data (1000 groups)
- **50** replicates → results are average estimates



# Parameters for simulated data

- **Simulated parameters:**
  - Direct variance ( $\sigma_{a_D}^2$ ) = 0.3
  - Indirect variance ( $\sigma_{a_I}^2$ ) = 0.003
  - Correlation between direct and indirect ( $r_{a_{DI}}$ ) = 0
- **Heritability** (direct and indirect): **0.3**
- **Estimation of genetic parameters:** DMU software



# Models

**Classical model (CM):**

$$y = Xb + Z_D a_D + Z_g g + e$$

Phenotype Sex (fixed) Direct Group Residual

**Indirect genetic model (IGM):**

$$y = Xb + Z_D a_D + Z_I a_I + Z_g g + e$$

Phenotype Sex (fixed) Direct Indirect Group Residual

# Predictive ability

The correlation between true phenotypes ( $P$ ) and estimated breeding values (EBVs) of individuals ( $i$ ) in the validation data

A.  $r_{P_i, DBV_{C_i}}$ :  $P_i$  correlated with **direct EBVs from CM**

B.  $r_{P_i, DBV_{I_i}}$ :  $P_i$  correlated with **direct EBVs from IGM**

C.  $r_{P_i, \sum_j^{n-1} IBV_{I_j}}$ :  $P_i$  correlated with **sum of indirect EBVs of group mates ( $j$ )**

D.  $r_{P_i, DBV_{I_i} + \sum_j^{n-1} IBV_{I_j}} = r_{total}$ :  $P_i$  correlated with **total breeding values (B+C)**

# Results – Estimated variance components

True parameters		0.3	0.003	0.0
Scheme	Model	$\sigma_{a_D}^2$ (SE)	$\sigma_{a_I}^2$ (SE)	$r_{a_{DI}}$ (SE)
1. dat <sub>full</sub>	IGM	0.323 (0.041)	0.003 (0.001)	-0.024 (0.173)
	CM	0.323 (0.041)	-	-
2. dat <sub>train</sub>	IGM	0.274 (0.036)	0.003 (0.001)	0.097 (0.182)
	CM	0.272 (0.036)	-	-

# Results – Predictive ability

Scheme	$r_{Phen_i, DBV_{C_i}}$	$r_{phen_i, DBV_{I_i}}$	$r_{phen_i, \sum_j^{n-1} IBV_{I_j}}$	$r_{total}$
1. dat <sub>full</sub>	0.276	0.276	0.040	0.279
2. dat <sub>train</sub>	0.272	0.271	0.038	0.274
3. dat <sub>valid</sub>	0.262	0.262	0.040	0.265
4. dat <sub>train_valid</sub>	0.258	0.258	0.041	0.261

# Conclusion

Phenotypic censoring does not affect the estimation of indirect genetic variance, nor prediction ability of indirect genetic effects

# Discussion

- More than one missing record per group
- Non-random censoring
- Average genetic relationship within groups other than random
- Overlapping generations or genetic relationships among animals in training and validation data