Stable indirect predictions with a large number of genotyped animals

Daniela Lourenco

Andre Garcia, Yutaka Masuda, Shogo Tsuruta, Ignacy Misztal
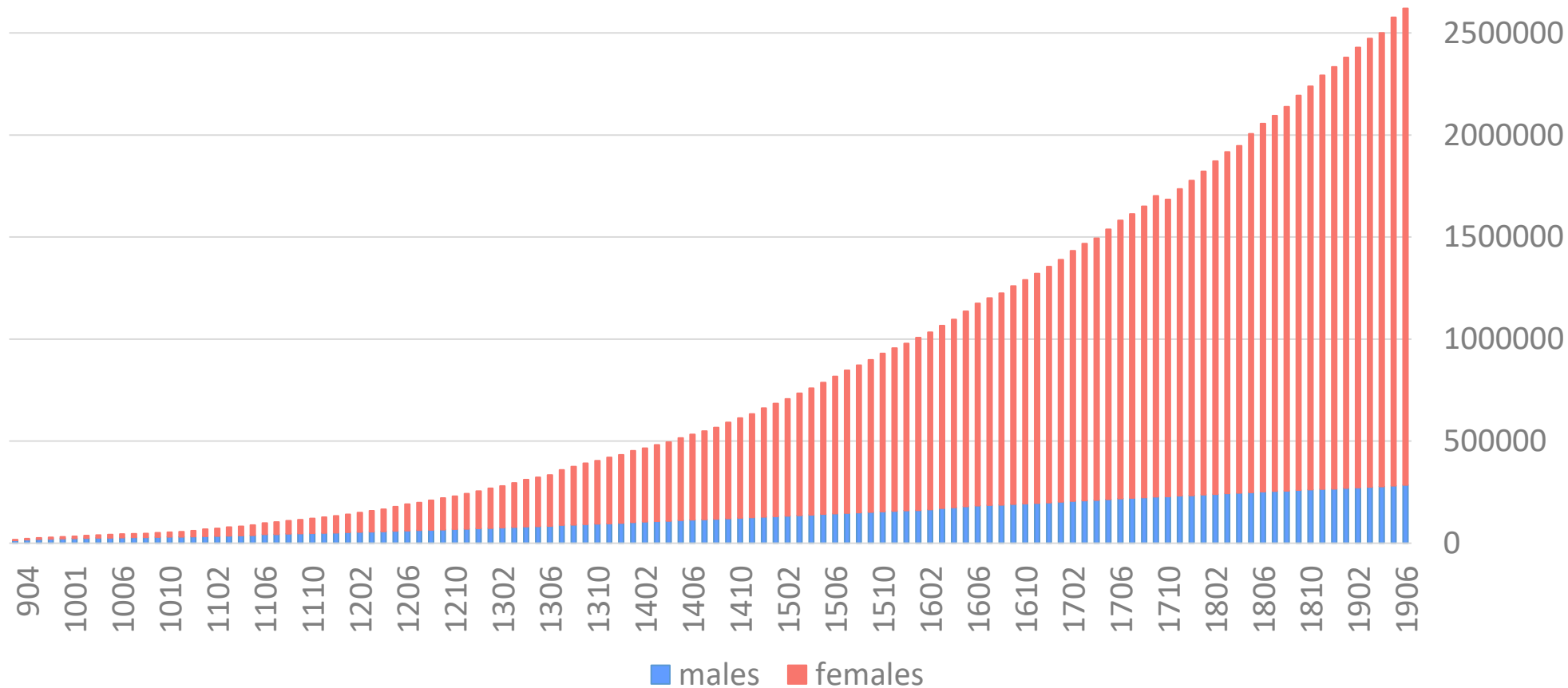
Interbull 2019 – Cincinnati, OH

Ever-increasing number of genotypes

US Holsteins

89% Females

20% Phenotypes

16% PED miss

■ males ■ females

https://queries.uscdcb.com/Genotype/cur_density.html

Do we need to include all genotyped animals in the evaluations?

Indirect Predictions

- GBLUP

- ssGBLUP

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'W} \\ \mathbf{W'X} & \mathbf{W'W+G^{-1}\lambda} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{W'y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'W} \\ \mathbf{W'X} & \mathbf{W'W+H^{-1}\lambda} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{W'y} \end{bmatrix}$$

SNP effects

GEBVs

$$\hat{a} = \lambda \mathbf{D} \, \mathbf{Z'G}^{-1} \hat{u}$$

VanRaden 2008
Stranden & Garrick 2009
Wang et al. 2012

$$\hat{u}_{ip} = \mathbf{Z}^{*} \hat{a}$$

When to use indirect predictions

- Not all genotyped animals are in the evaluations
  - Animals with incomplete pedigree increase bias and lower R^2

- Interim evaluations
  - Between official runs

- Commercial products
  - e.g. GeneMax for non-registered animals
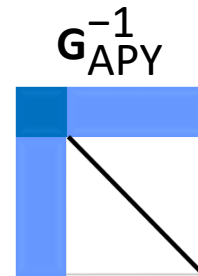
APY and Indirect Predictions

$$\hat{a} = \lambda \mathbf{D} \ \mathbf{Z}' \mathbf{G}^{-1} \hat{u}$$

G⁻¹

$$\mathbf{G}_{APY}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{cc}^{-1}\mathbf{G}_{cn} \\ I \end{bmatrix} \mathbf{M}_{nn}^{-1} \begin{bmatrix} -\mathbf{G}_{nc}\mathbf{G}_{cc}^{-1} & I \end{bmatrix}$$

Misztal et al., 2014

$$\hat{a} = \lambda \mathbf{D} \ \mathbf{Z}' \mathbf{G}_{APY}^{-1} \hat{u}$$

$\mathbf{G}_{APY}^{-1}$

CORR(IP$_1$,IP$_2$) = 0.99

Lourenco et al., 2018

Common practice in APY

- Select core animals

    - Randomly

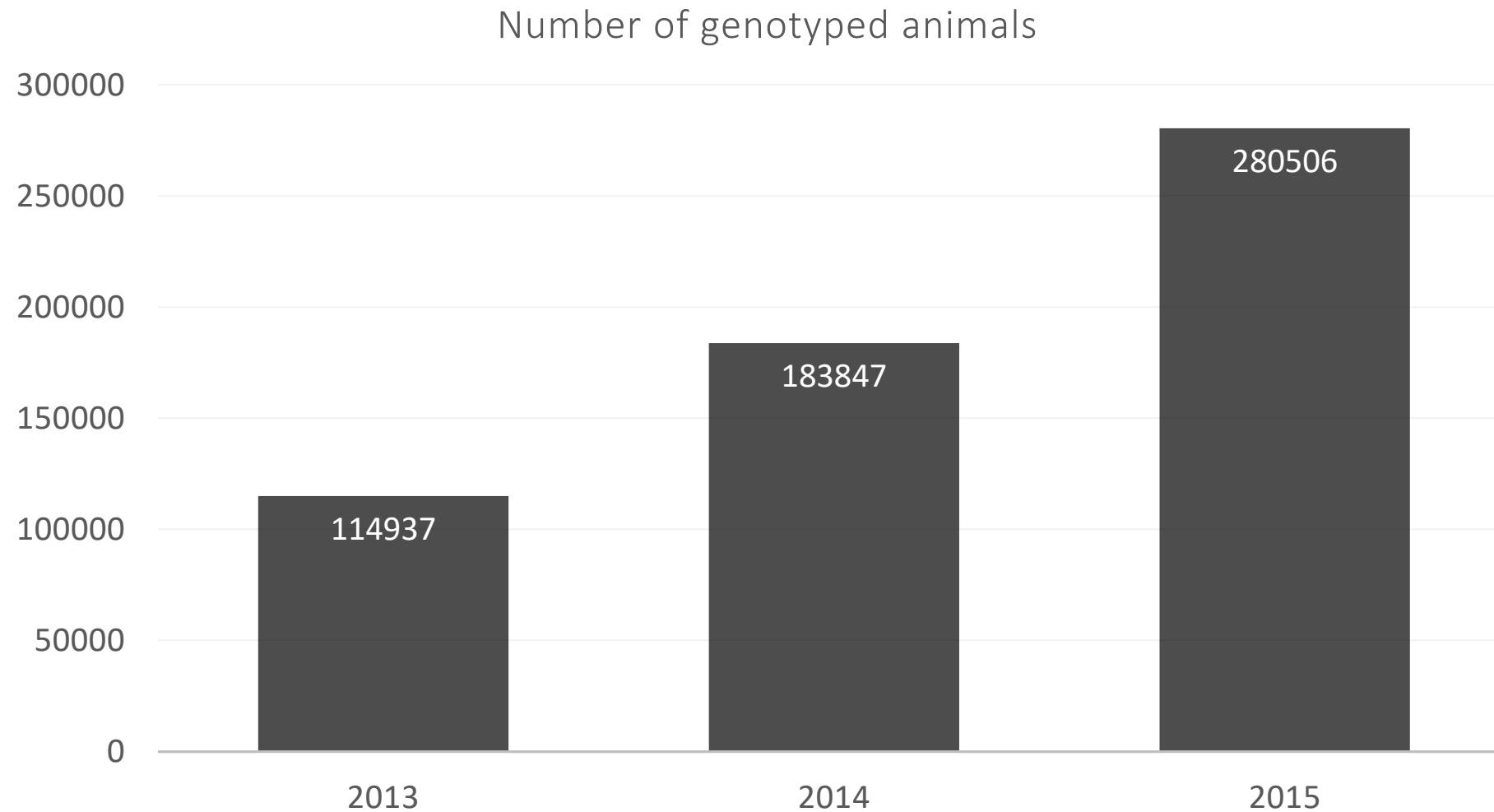   - Amount of information

- Keep the same core for several runs

What happens with IP when the number of genotyped animals increases under APY?

Data

- American Angus Association
- BW, **WW**, PWG

Andre Garcia

Number of genotyped animals

Year	Count
2013	114937
2014	183847
2015	280506

Selecting 19k Core Animals

$$\mathbf{G}_{\text{APY}}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{cc}^{-1}\mathbf{G}_{cn} \\ \mathbf{I} \end{bmatrix} \mathbf{M}_{nn}^{-1} [-\mathbf{G}_{nc}\mathbf{G}_{cc}^{-1} \quad \mathbf{I}]$$

- Fixed core
  - Chosen randomly in 2013

- Updated core
  - Chosen randomly every year

- Extra scenarios
  - Oldest - born up to 2010
  - Parents - born up to 2013
  - Youngest - born in 2015

GBLUP: $\widehat{\boldsymbol{u}}_{ip} = \mathbf{Z}\hat{a}$

ssGBLUP: $\widehat{\boldsymbol{u}}_{ip} = \overline{GEBV_{eval}} + \mathbf{Z}\hat{a}$

- GEBV $(\widehat{\boldsymbol{u}})$
- SNP effects

  $\hat{\boldsymbol{a}} = \lambda \mathbf{D}\,\mathbf{Z}'\mathbf{G}_{\text{APY}}^{-1}\widehat{\boldsymbol{u}}$

- IP for all animals

- CORR $(\widehat{\boldsymbol{u}},\ \widehat{\boldsymbol{u}}_{ip})$

Correlation between GEBV and IP

- SNP effects

$$\hat{a} = \lambda \mathbf{D} \, \mathbf{Z}' \mathbf{G}_{\mathrm{APY}}^{-1} \hat{u}$$
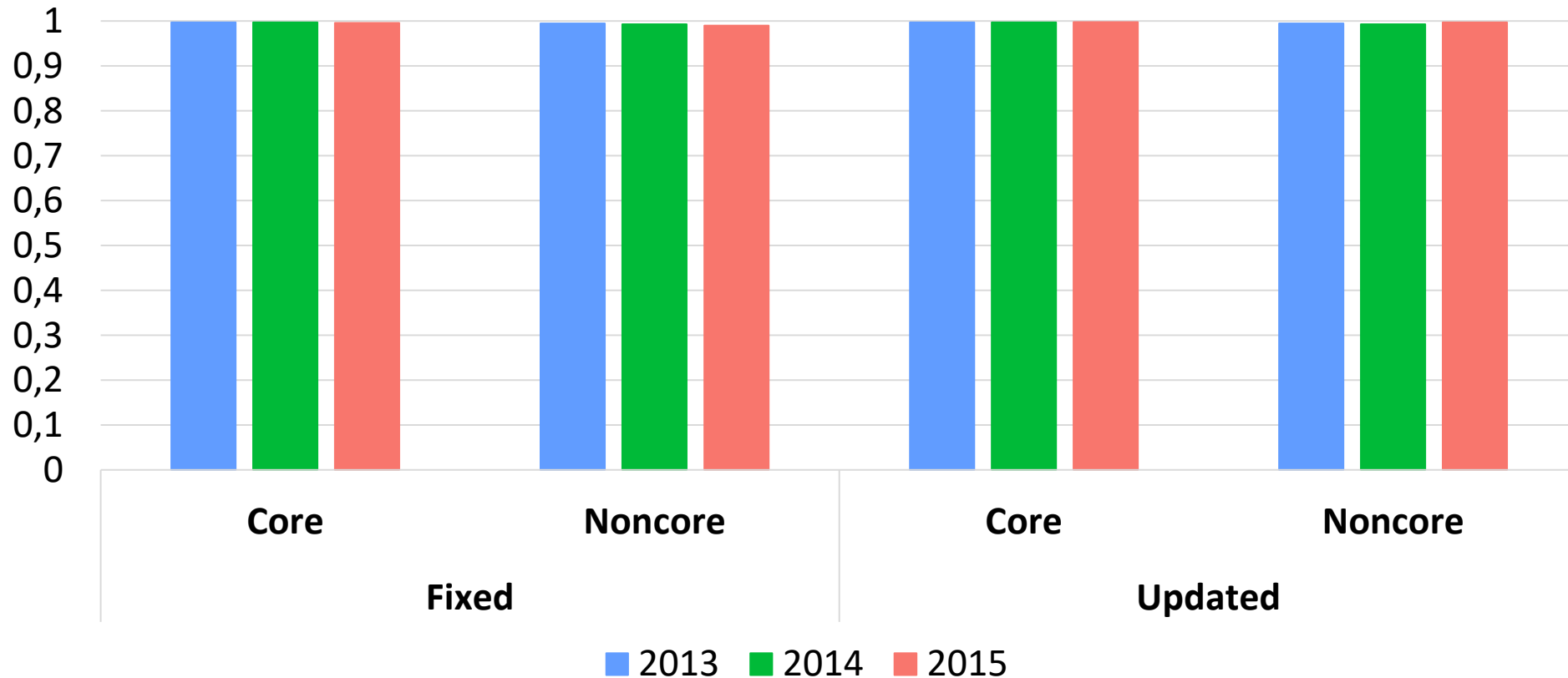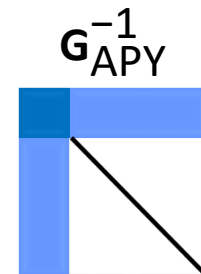
APY and Indirect Predictions

$$\mathbf{G}_{\text{APY}}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{cc}^{-1}\mathbf{G}_{cn} \\ I \end{bmatrix} \mathbf{M}_{nn}^{-1} \begin{bmatrix} -\mathbf{G}_{nc}\mathbf{G}_{cc}^{-1} & I \end{bmatrix}$$

Misztal et al., 2014

$$\hat{a} = \lambda \mathbf{D}\, \mathbf{Z}'\mathbf{G}^{-1}\hat{u}$$

$\mathbf{G}^{-1}$

$$\hat{a} = \lambda \mathbf{D}\, \mathbf{Z}'\mathbf{G}_{\text{APY}}^{-1}\hat{u}$$

$\mathbf{G}_{\text{APY}}^{-1}$

CORR(IP$_1$,IP$_2$) = 0.99

$$\hat{a} = \lambda \mathbf{D}\, \mathbf{Z}'\mathbf{G}_{\text{Core}}^{-1}\hat{u}_{\text{Core}}$$

$\mathbf{G}^{-1}$ core

CORR(IP$_1$,IP$_3$) = 0.99

Lourenco et al., 2018

What happens with IP when the number of genotyped animals increases under APY and SNP effects are computed based only on CORE animals?

Correlation between GEBV and IP

- SNP effects

$$\widehat{a} = \lambda \mathbf{D} \, \mathbf{Z}' \mathbf{G}_{\text{Core}}^{-1} \widehat{u}_{\text{Core}}$$
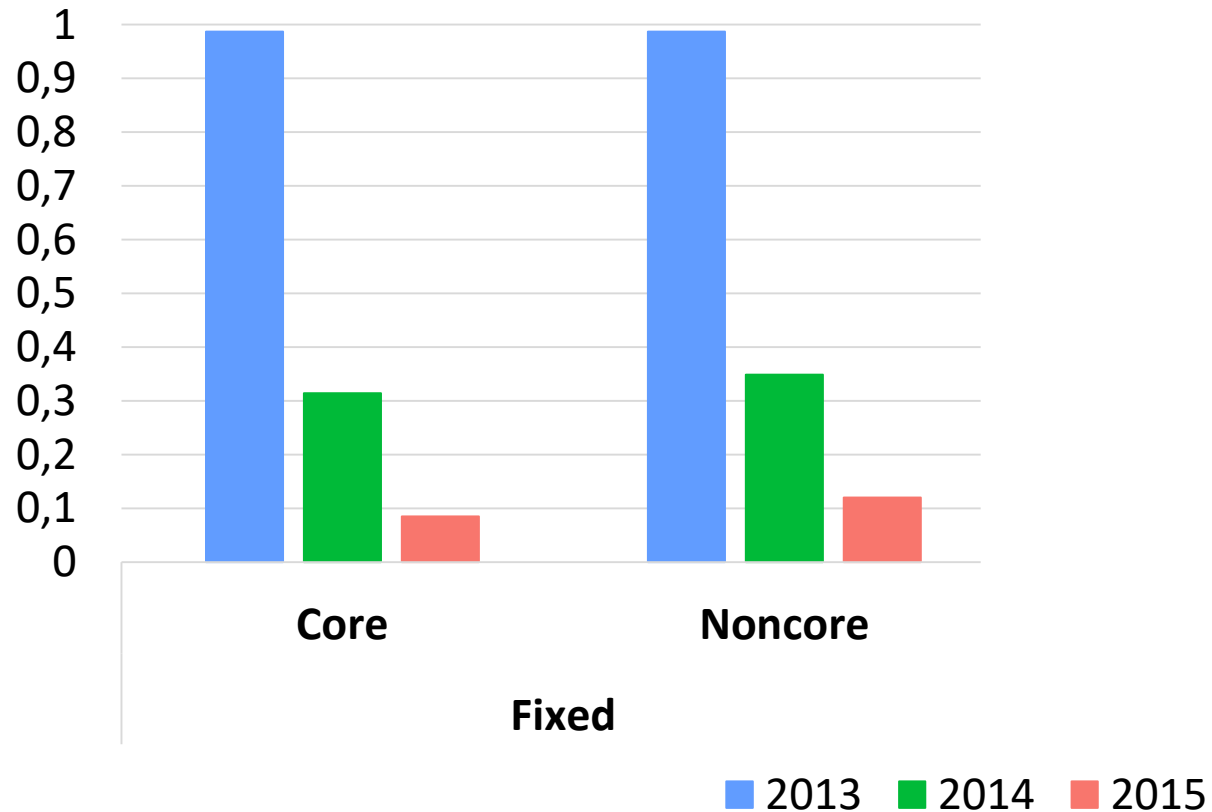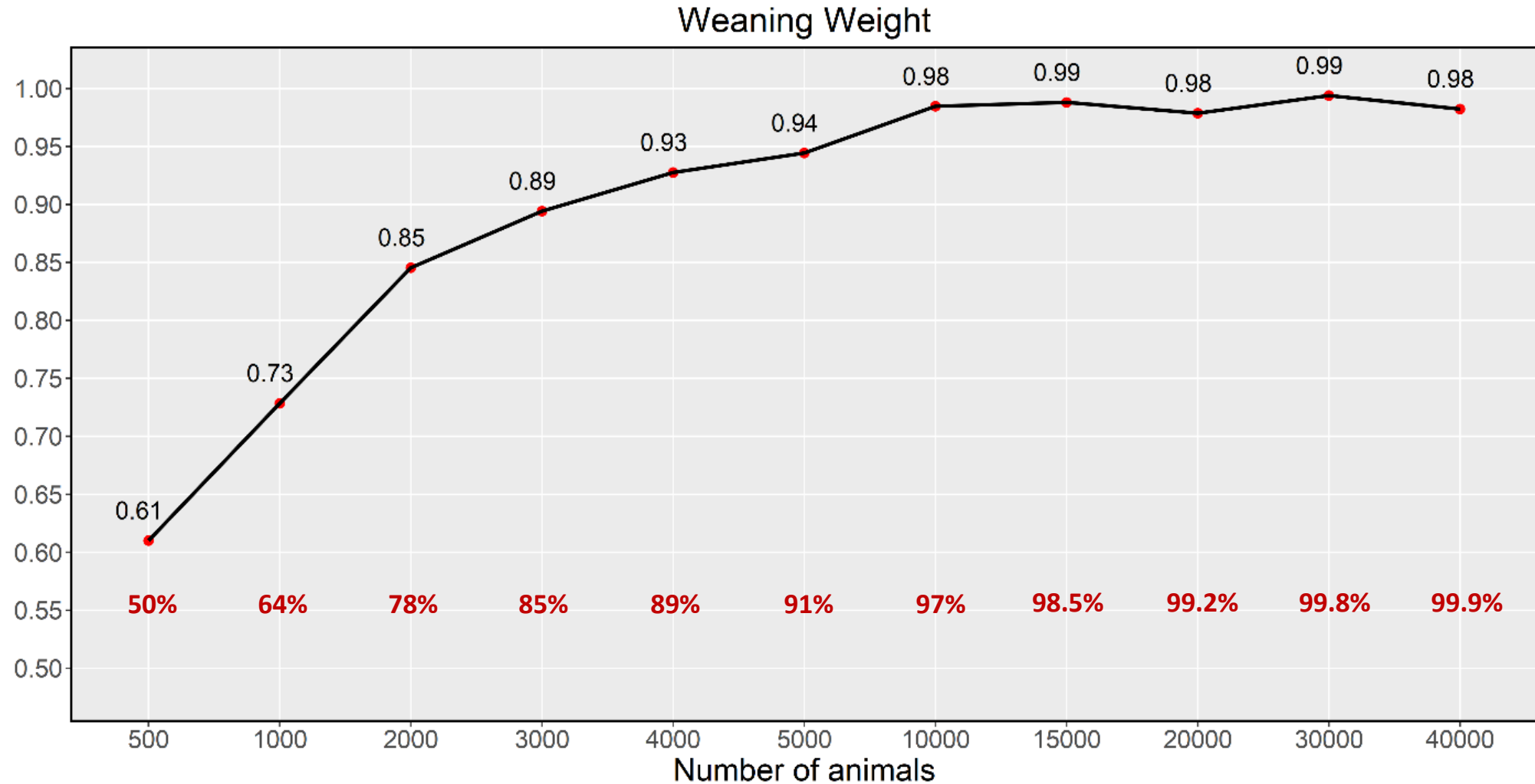
Reverse Engineering

- Only a small portion of the data
  - Approximate SNP effects
  - Predictions for genotyped animals

What happens with predictions under reverse engineering?

Correlation between GEBV and Predictions

- Data up to 2015

Weaning Weight

Final Remarks

- **If all genotyped animals are used to compute SNP effects**

  - Indirect Predictions are robust

  - Independent of core choice

- **If only core animals are used**

  - Robust Indirect Predictions with updated core

  - Core should reflect the dimensionality of **G** (98% - 99%)

- **If only a small portion of the data is available**

  - Predictions are less accurate

Acknowledgements

Steve Miller

Dan Moser