

64th Annual Meeting of the EAAP 2013
Nantes, 27 August 2013

Validation accuracy of genomic breeding values with HD genotypes in Fleckvieh cattle

J. Ertl,* C. Edel,* R. Emmerling,* H. Pausch,† R. Fries,† K.-U. Götz*

*Institute of Animal Breeding, Bavarian State Research Centre for Agriculture,
85586 Poing, Germany

† Chair of Animal Breeding, Technische Universität München, 85354 Freising,
Germany

Introduction

- Genomic Selection with GBLUP (VanRaden, 2008)
 - ◆ Realized relationships estimated with markers
 - ◆ More reliable BV than with pedigree relationships

- Little or no advantage in validation reliability with HD genotypes as compared to 50K (in Holstein)
 - ◆ With GBLUP (Erbe et al., 2012; Su et al., 2012)
 - ◆ But neither with Bayesian methods (Harris et al., 2011; Erbe et al., 2012; Su et al., 2012; VanRaden et al., 2013)

Objectives

- Advantage of HD genotypes in Fleckvieh?
- Is the advantage significant?
- Impacts of HD on model based reliability and inflation?

Dataset

- 10,240 Fleckvieh bulls
 - ◆ 1,492 HD genotyped
 - ◆ 8,748 50K genotyped, HD imputed (FImpute)
- Aggregated phenotypes
 - ◆ DYD in milk yield, fat yield, protein yield, SCS, muscling, udder, feet and legs, stature
 - ◆ DRP in milkability

Forward Prediction

□ Reference / Validation (split date 1.4.2005)

□ GBLUP

◆ $\mathbf{G}^* = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum_{k=1}^m p_k(1-p_k)}$ (VanRaden, 2008)

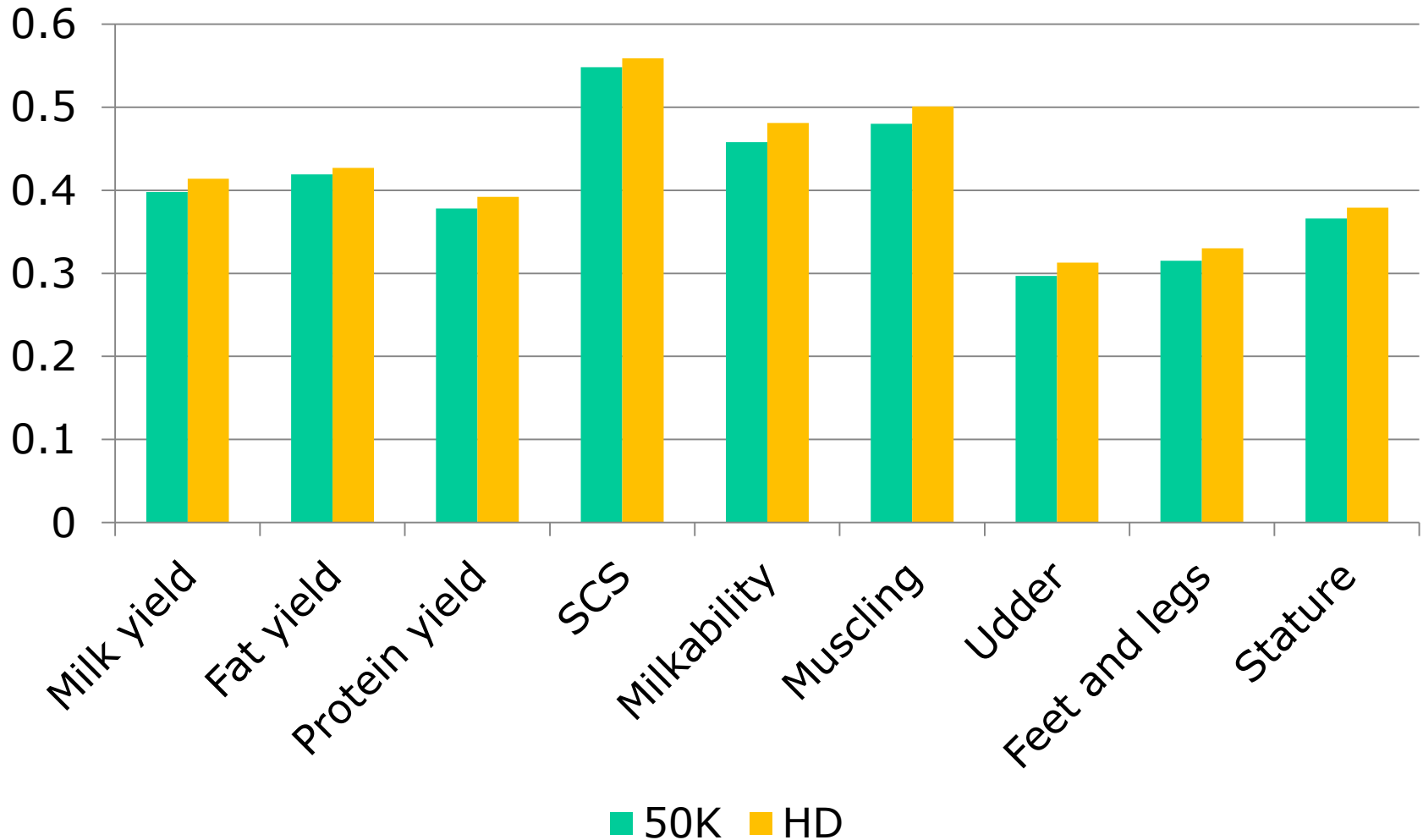
◆ \mathbf{G}^* scaled to **NRM** (Meuwissen et al., 2011)

◆ $\mathbf{G} = 0.99 \mathbf{G}^* + 0.01 \mathbf{NRM}$

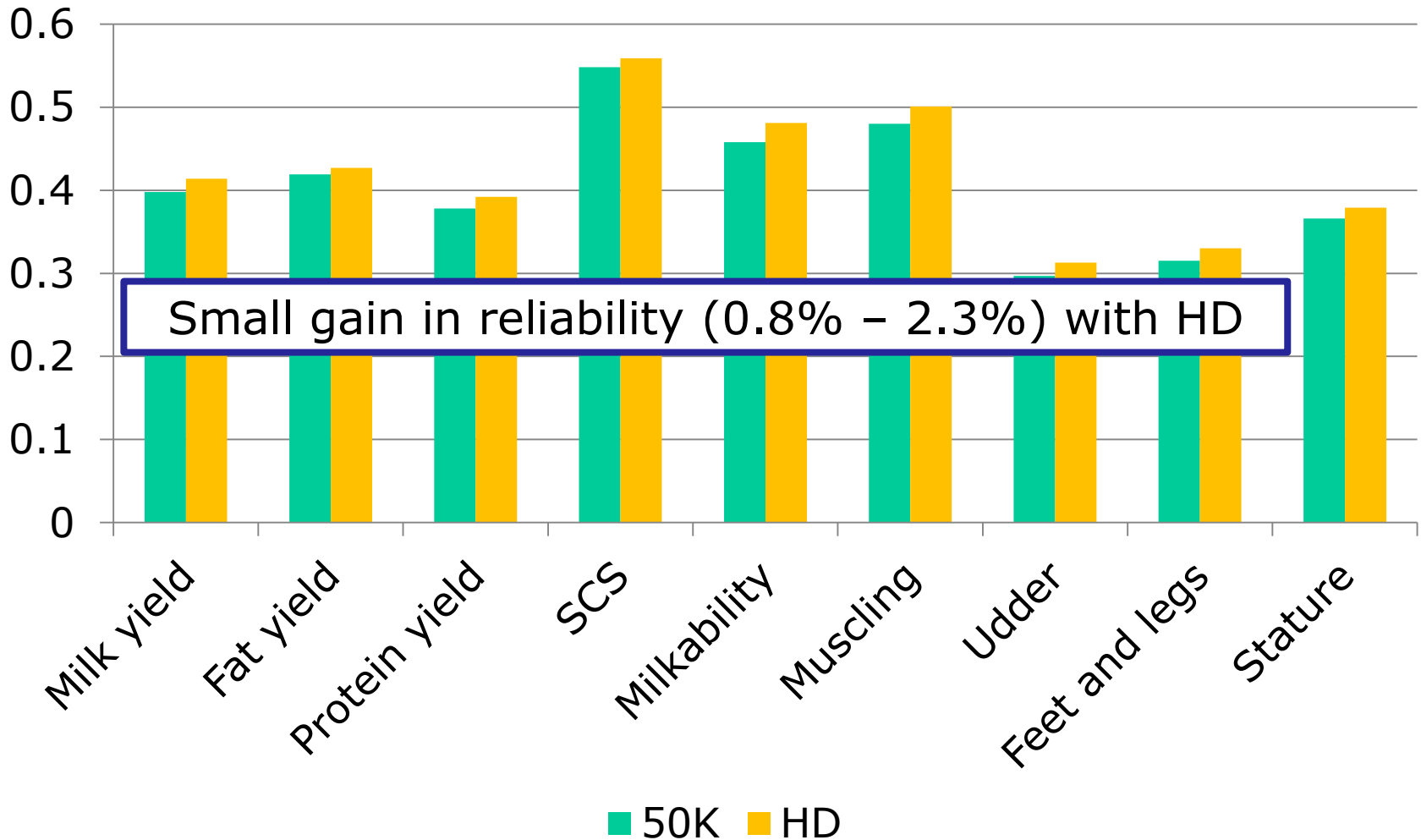
□ Validation reliability: $\frac{R^2[\text{regression}(\mathbf{D}\mathbf{Y}\mathbf{D}, \hat{\mathbf{g}})]}{R_{D\mathbf{Y}\mathbf{D}}^2}$

□ Inflation: $b(\mathbf{D}\mathbf{Y}\mathbf{D}, \hat{\mathbf{g}})$

Validation reliability



Validation reliability



Validation reliability

- ❑ Slightly larger validation reliability with HD genotypes (difference: 0.8% – 2.3%)
- ❑ Comparable with results in Nordic Red cattle (Su et al., 2012)
- ❑ Is this gain in validation reliability significant?

Distribution of 50K validation reliability?

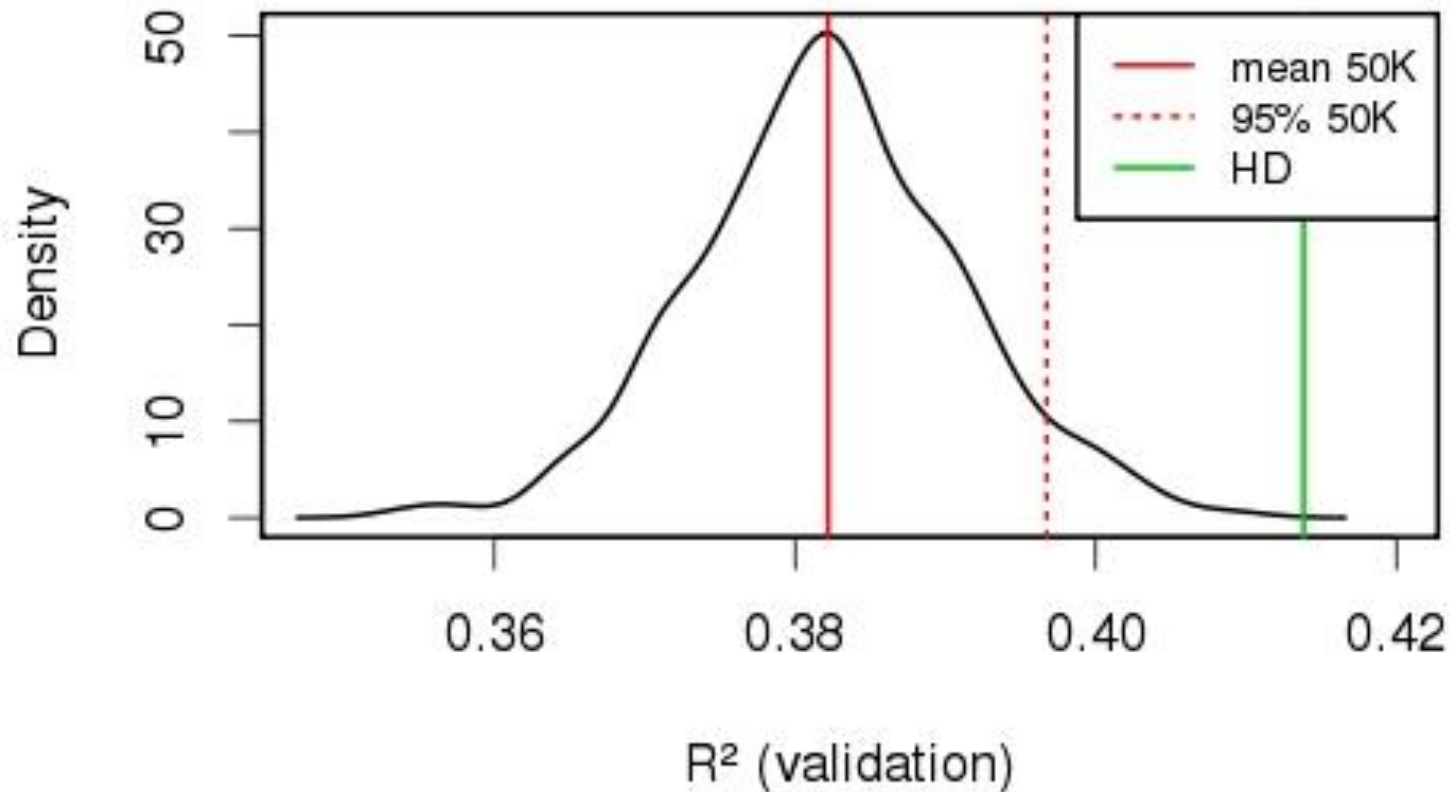
- ❑ Observed validation reliability with 50K chip
- ❑ What is the distribution?
- ➔ Repeated sampling of 50K SNP out of HD ($n=500$)
 - ◆ Stratified samples with structure similar to real 50K
- ❑ Forward prediction and validation with each 50K sample
- ➔ Distribution of 50K validation reliability

Is reliability gain significant?

- ❑ Comparison of HD reliability with 50K distribution
- ❑ „fair“ comparison because
 - ◆ Same level of imputing error for both SNP densities
 - ◆ Represents the situation that bulls are genotyped with both chips or that imputation is possible without error
- ❑ HD is significantly better if:
 $R^2_{HD} > 95\%$ quantile of R^2_{50K} samples

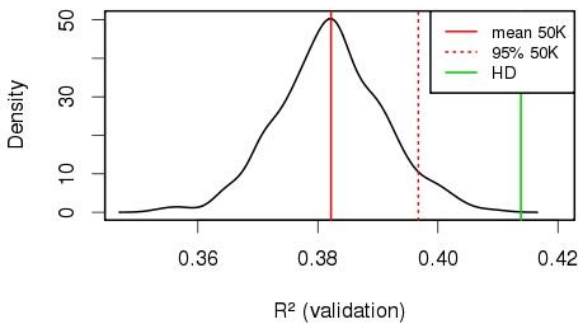
Distribution of 50K validation reliability

Milk yield

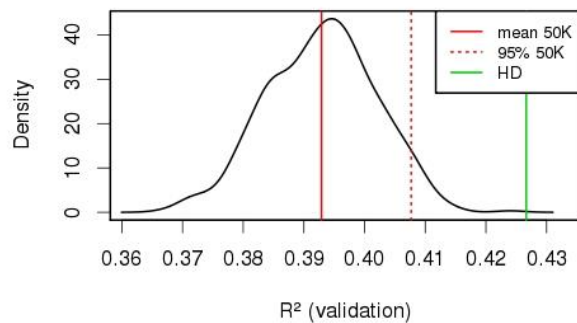


Distribution of 50K validation reliability

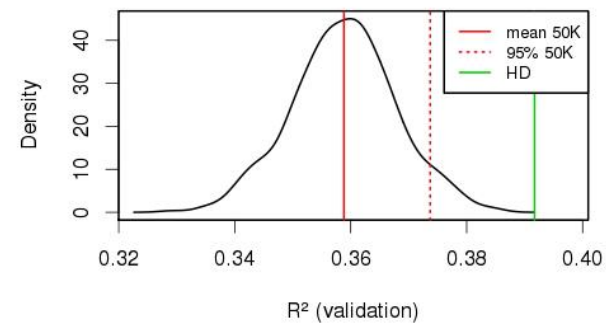
Milk yield



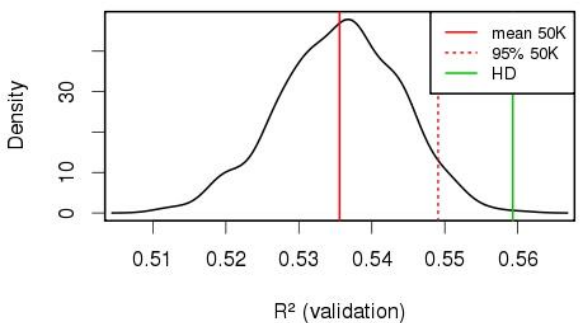
Fat yield



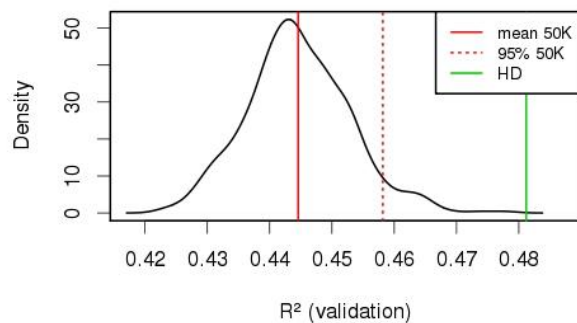
Protein yield



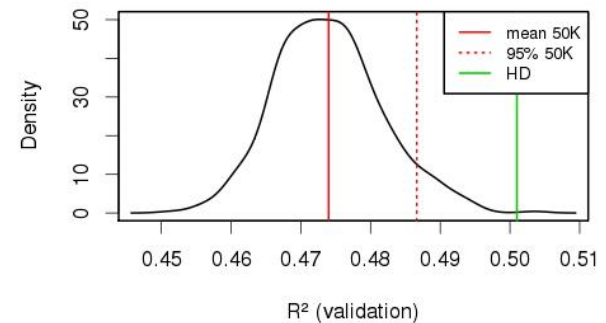
Somatic cell score



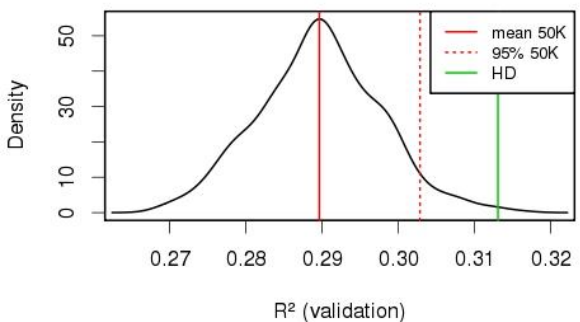
Milkability



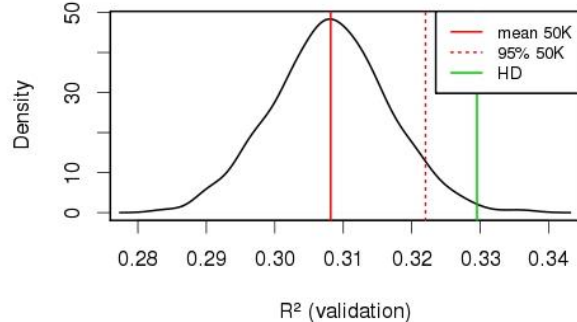
Muscling



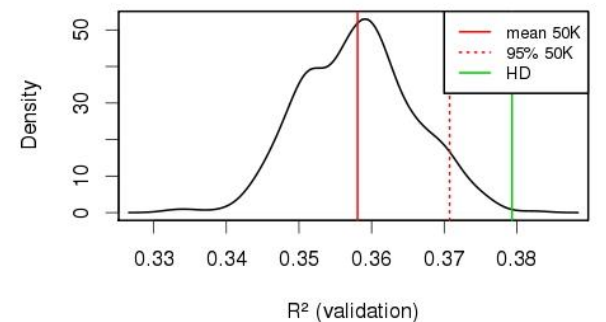
Udder



Feet and legs



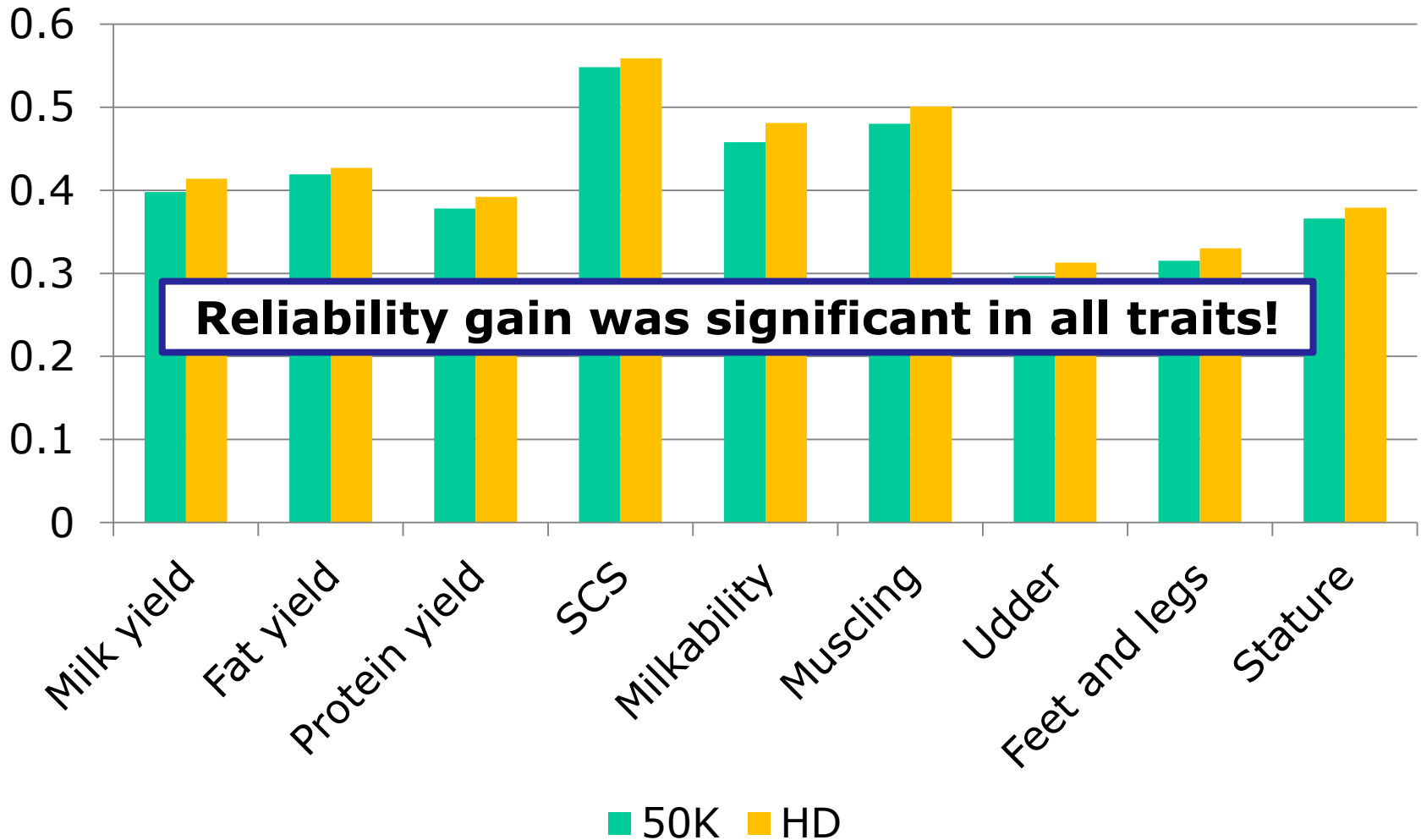
Stature



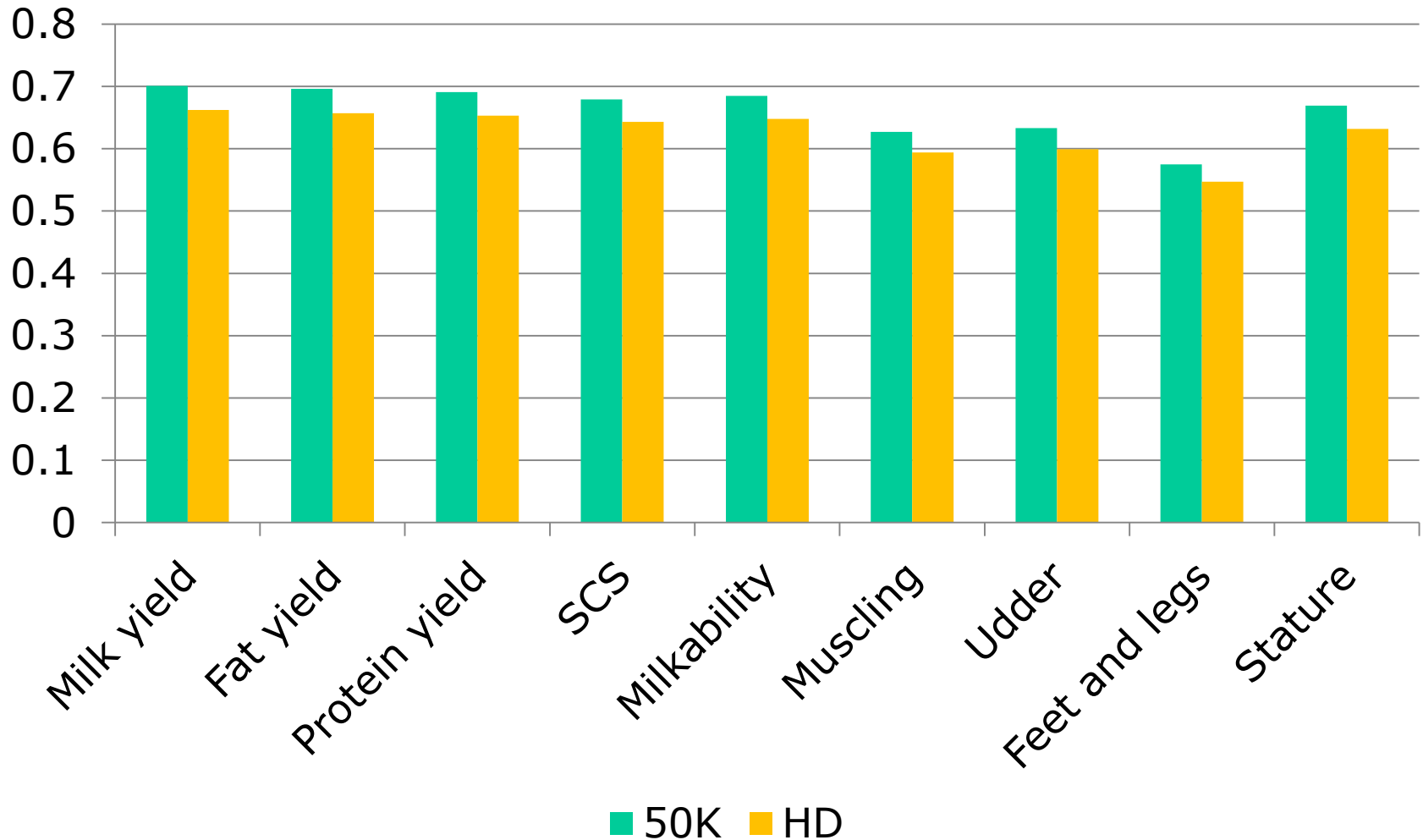
Varying the validation group

- ❑ So far, results rely on 1 validation group
- ➔ Repeated sampling of 500 validation animals with replacement (fixed calibration group)
- ➔ Validation reliability for 50K and HD
- ❑ „relevant“ comparison:
 - ◆ 50K are actually genotyped, no imputing error
 - ◆ HD contain imputing error
 - ◆ Represents current situation that all candidate and AI bulls are genotyped for 50K, but only a fraction for HD
- ❑ HD advantage: **2.1% - 3.6%**
- ❑ One-sided paired t-test: **$p < 0.001$**

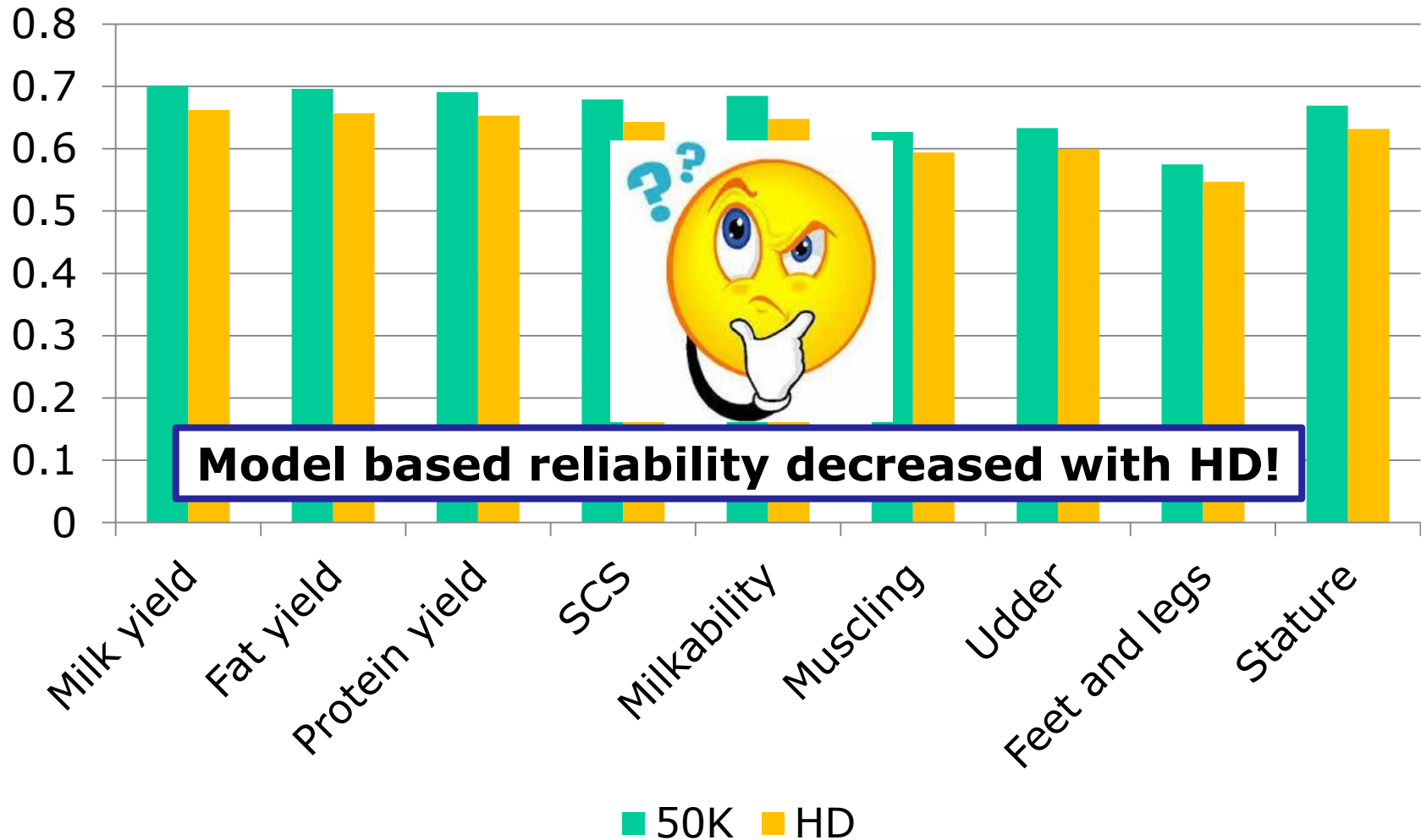
Validation reliability



Model based reliability



Model based reliability



Reason for decreased model based reliability?

- ❑ Sampling error of genomic relationships causes overestimation of model based reliability (Goddard et al., 2011)
- ❑ Sampling error of genomic relationship coefficients:

$$\frac{n}{M \cdot CV^2} \quad (\text{Endelman \& Jannink, 2012})$$

- ➡ Sampling error (50K) >> sampling error (HD)
- ➡ Model based reliability (HD) is closer to the truth

2 ways to reduce sampling error

1. Higher marker density (HD vs. 50K)

◆ Sampling error ↓

◆ More detailed genomic information → validation R^2 ↑

2 ways to reduce sampling error

1. Higher marker density (HD vs. 50K)

- ◆ Sampling error ↓

- ◆ More detailed genomic information → validation R^2 ↑

2. Shrinkage estimation

- ◆ Shrinkage of **G** towards diagonal matrix

(Endelman & Jannink, 2012)

- ◆ Shrinkage intensity (50K) = 2% $\frac{n}{M \cdot CV^2}$

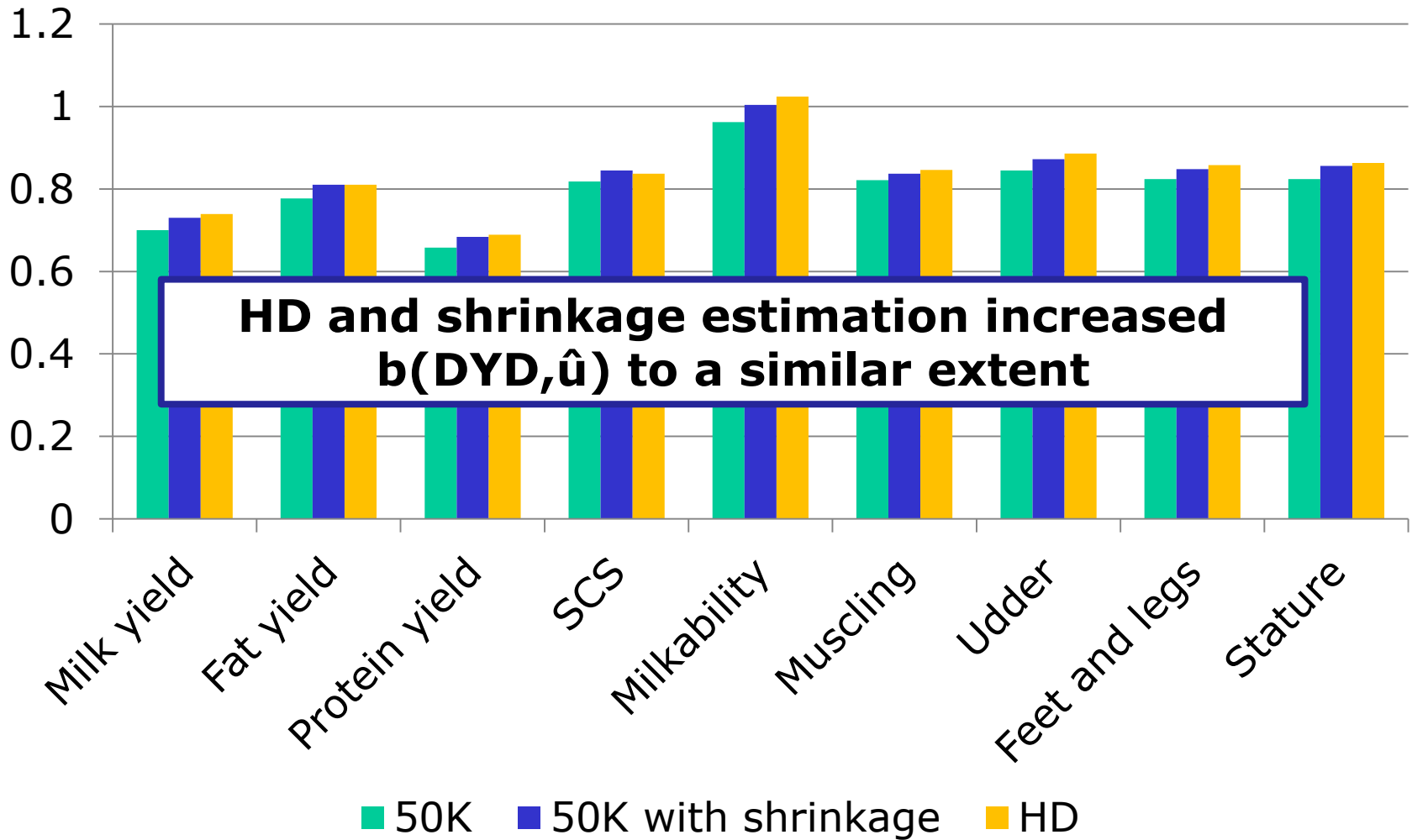
- ◆ Correction of overdispersion of genomic relationships

→ model based R^2 ↓

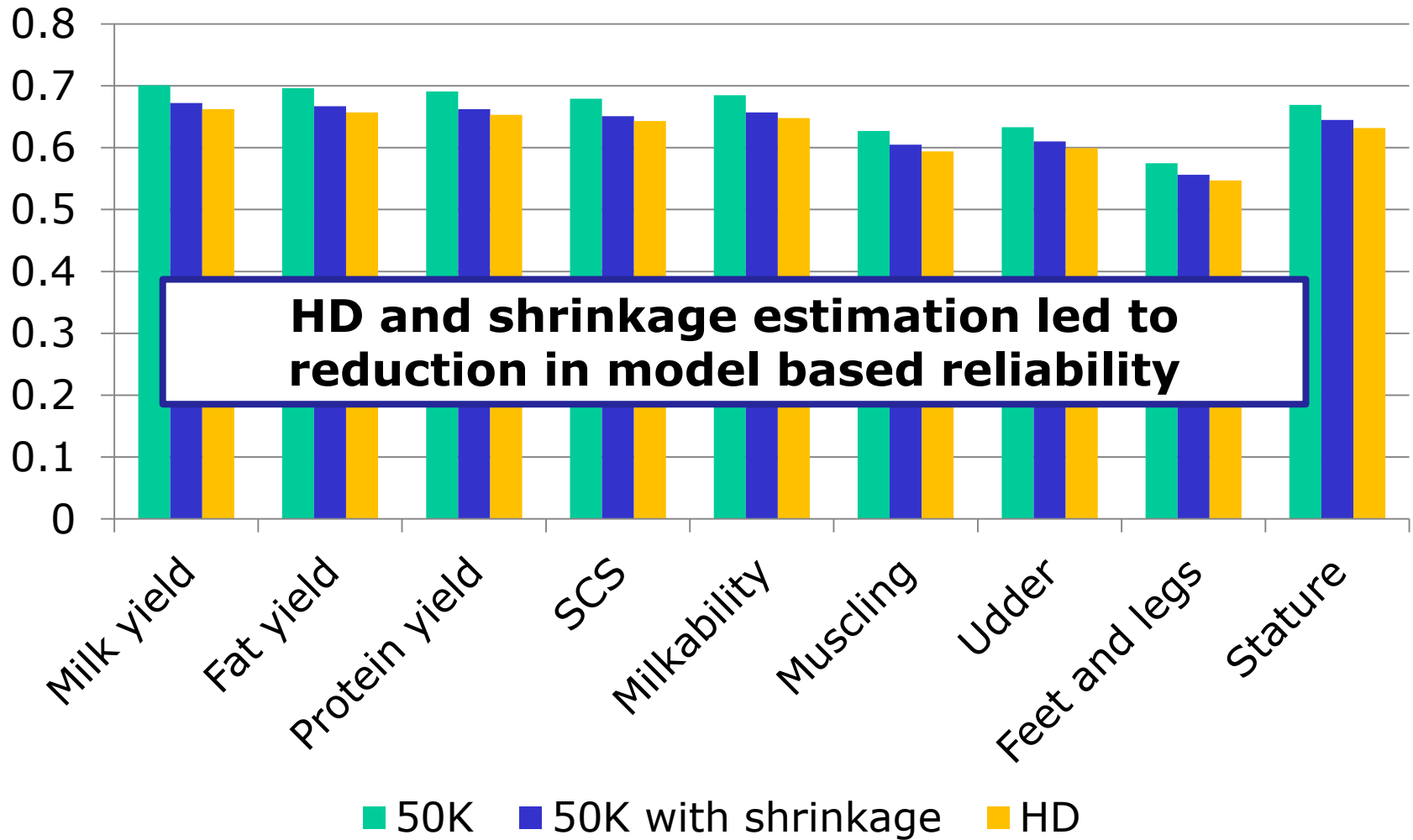
→ inflation ↓

- ◆ **But:** no additional genomic information

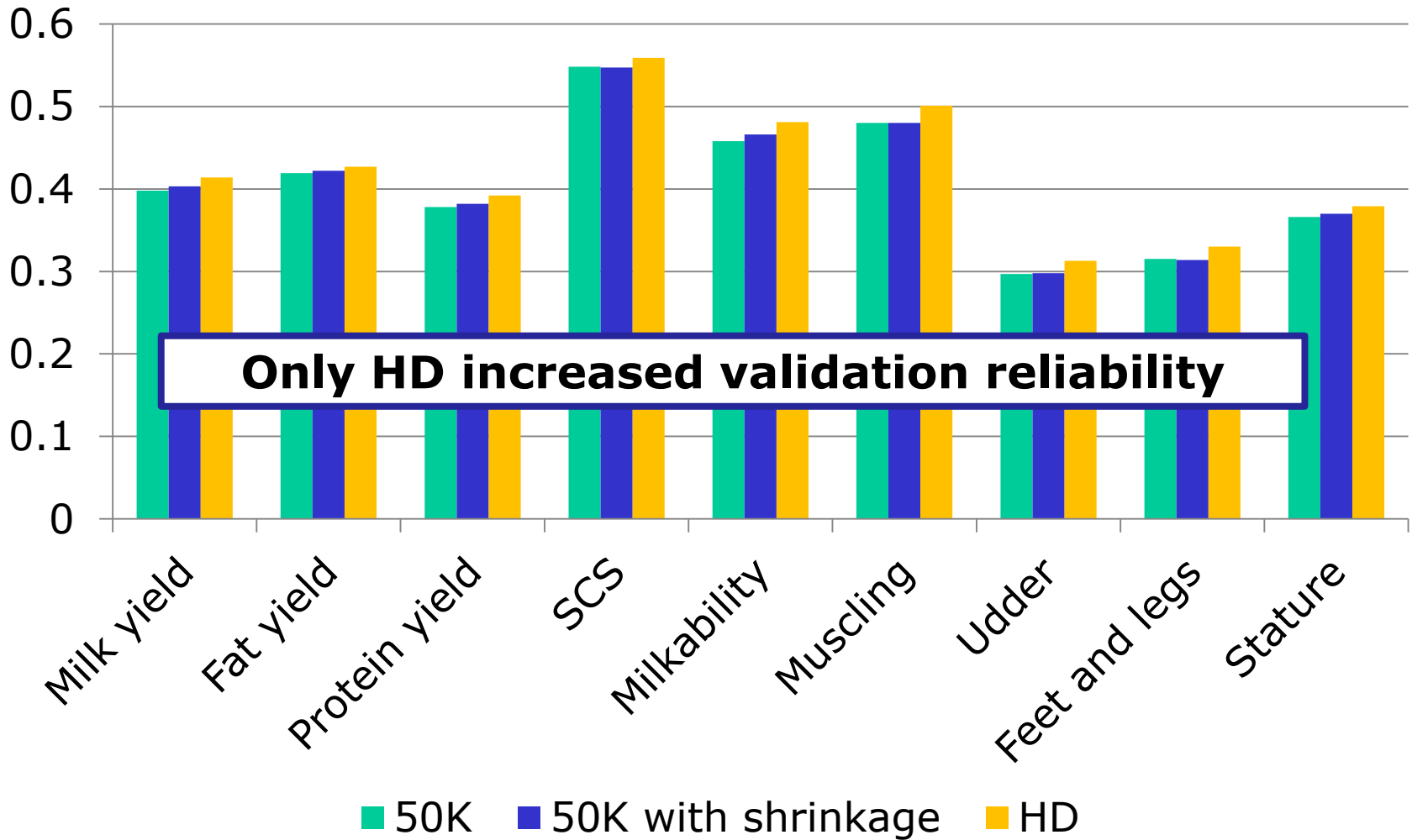
Regression coefficients



Model based reliability



Validation reliability



Conclusions

- ❑ Small gains in validation reliability with HD (also in Fleckvieh)
 - ◆ Relevant comparison: 0.8% – 2.3%
 - ◆ Fair comparison: 2.1% – 3.6%
 - ◆ Statistically significant
- ❑ Model based reliability: less overestimation with HD
- ❑ Less inflation with HD
- ❑ Other benefits with HD:
 - ◆ Phased genotypes available
 - ◆ Imputation to sequence
 - ◆ Hereditary diseases
- ❑ Recommendation to HD genotype AI bulls



Bavarian State Ministry of Food, Agriculture and Forestry



SPONSORED BY THE



Federal Ministry of Education and Research

This research was funded by the German Federal Ministry of Education and Research (BMBF) within the AgroClustEr „Synbreed – Synergistic plant and animal breeding“ (Grant ID 0315528H). We want to thank M. Sargolzaei for providing FImpute software.

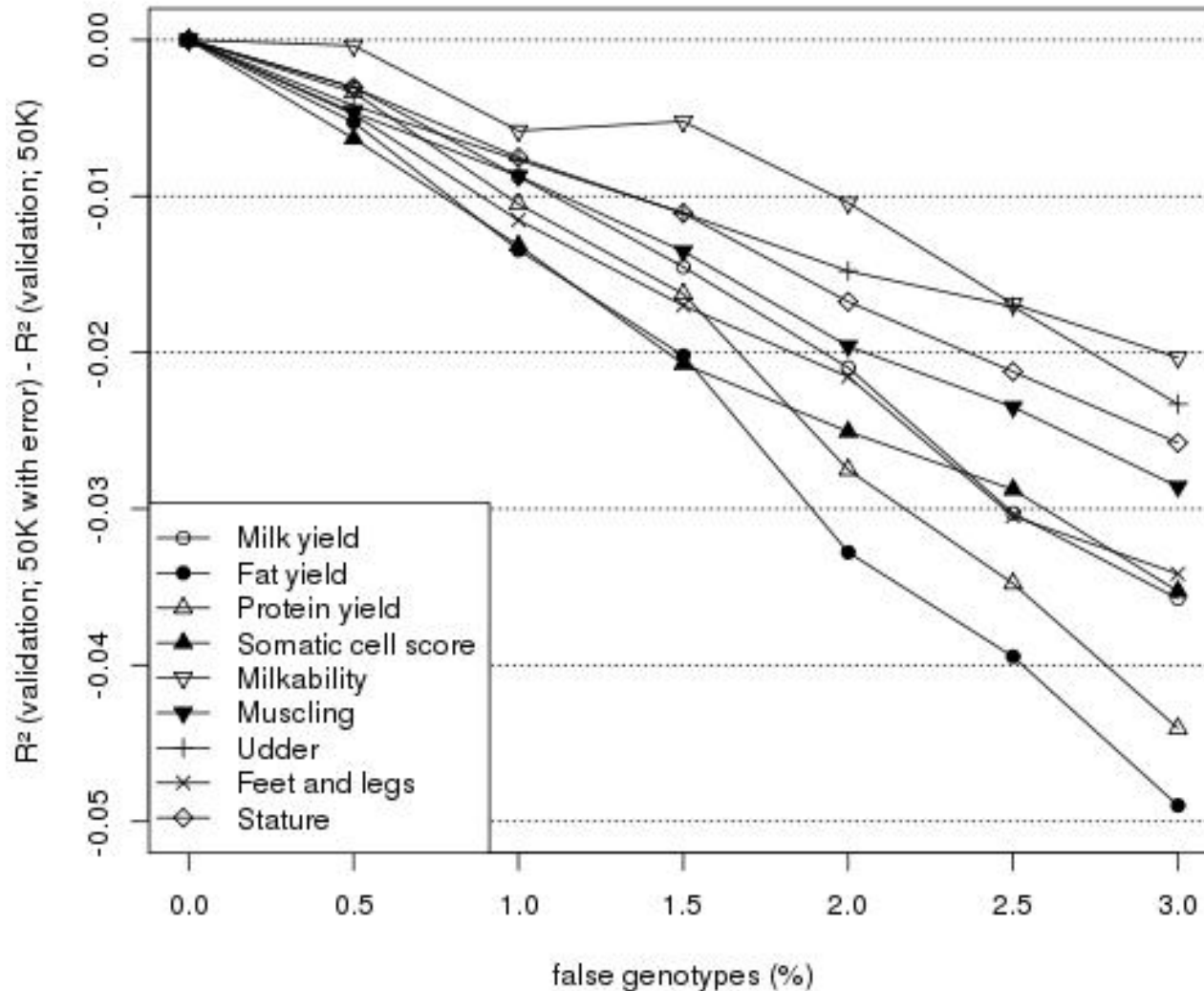
Thank you for your attention!



Christian-Albrechts-Universität zu Kiel



Impact of imputing error on validation reliability



Model based reliability - example

