Genomic reliabilities

Testing the protocol proposed by Interbull

H. Eding, J. Vandenplas, G. de Jong

Dubrovnik, August 26, 2018

OVERVIEW

- Introduction

- Six step procedure

- Test data and parameters used

- Some results

- Discussion

INTRODUCTION

- Interbull started a working group Genomic Reliabilities

- Aim: Find a procedure to estimate animal genomics reliabilities
    - For multi- and single-step procedures
    - For genotyped and non-genotyped animals
    - Accounts for residual polygenic effects
    - Consistent results across countries (unified approach)
    - Feasible for large numbers if genotyped animals
    - Usable in routine genomic evaluations

INTRODUCTION

- AEU started development of a (D)PCG to solve singe step models
  - Single step GBLUP
  - Single step SNP BLUP
  - Aimed for use in the national (genomic) evaluation
    - Replaces our current pseudo-trait methods
- In parallel new software to estimate GEBV reliabilities is needed
  - Incorporates traditional sources of information (parents, offspring, own)
  - Incorporates information from genotypes
  - Accounts for propagation (effect of genomic information on offspring, parents)
  - Proposal Interbull working group chosen as approach.

CRV

BETTER COWS | BETTER LIFE

INTRODUCTION

- Project: Develop software/workflow
    - Implementing the Interbull protocol
    - Within frame work of new (D)PCG to solve single step SNP BLUP models
    - Produce correct GEBV reliabilities for genotyped and non-genotyped animals

- First test on existing flow: Milking speed and Temperament
    - To get a feel for the requirements
    - Compare results with current GEBV reliability estimates

SIX STEPS

1. Calculate SNP reliabilities
2. Derive DGV reliabilities
3. Adjust DGV reliabilities
4. Calculate genomic gain of reliabilities
5. Propagation (optional; non-gentyped animals)
6. Calculate final reliabilities

CRV

BETTER COWS | BETTER LIFE

GETTING PARAMETERS

- Rate of imputation
  - 'accuracy of genotype imputation'
  - All bulls on same chip, so r_{imp} = 0.985 (empirical mean)

- Theoretical to realized reliability factor f
  - Set to 1.d0 for first testing
  - Unclear how to derive correct value

- Proportion of residual polygenic variance k
  - Expected values ~ [0.05 ; 0.25]

GETTING PARAMETERS: K

- Testing on cow reference

$$\sigma^2_{SNP_{tot}} = \left(\left(\frac{N - \sum_j pp_j}{100} \right) + \sum_j pp_j \right) \sigma^2_{SNP}$$

$$k = \sigma^2_{poly} / (\sigma^2_{SNP_{tot}} + \sigma^2_{poly})$$

- Variance components from current genomic selection flows

CRV

BETTER COWS | BETTER LIFE

FIRST TEST: TEMPERAMENT AND MILKING SPEED

- Animal ID's
  - Pedigree   : 6,300,989 animals
  - Observations   : 4,021,612

- Genotypes
  - Most recent run (2018 05 01)
  - Number of genotypes   : 115,378
  - Number of SNP   :   37,995
  - Converted to [012]-format

CRV

BETTER COWS | BETTER LIFE

TEST DATA USED: TEMPERAMENT AND MILKING SPEED

- During genomic selection validation
  - Added EDC (Δ_{EDC}) are calculated from DGV and BLUP results
  - Based on differences in reliability EBV and GEBV for young bulls (no daughters)
  - Assumed constant (single genotype in single animal)

- Overview traits:

Trait	h^2	k	EDC$_{add}$ (val 2018)	Mean rel. (GEBV YB)
Milking Speed	0.230	0.084	24.1	0.71
Temperament	0.114	0.208	5.1	0.38

CRV

BETTER COWS | BETTER LIFE

RESULTS

RESULTS: RUNTIME

Step	Action	Time	
1	Conventional rels	3m10	
2	Read genotypes	1m20	
3	Convert to [012]	3m32	
4	Run luke software	53m20	per trait
5	Apply Liu protocol	2m00	

- Peak memory usage: 52.4 Gb
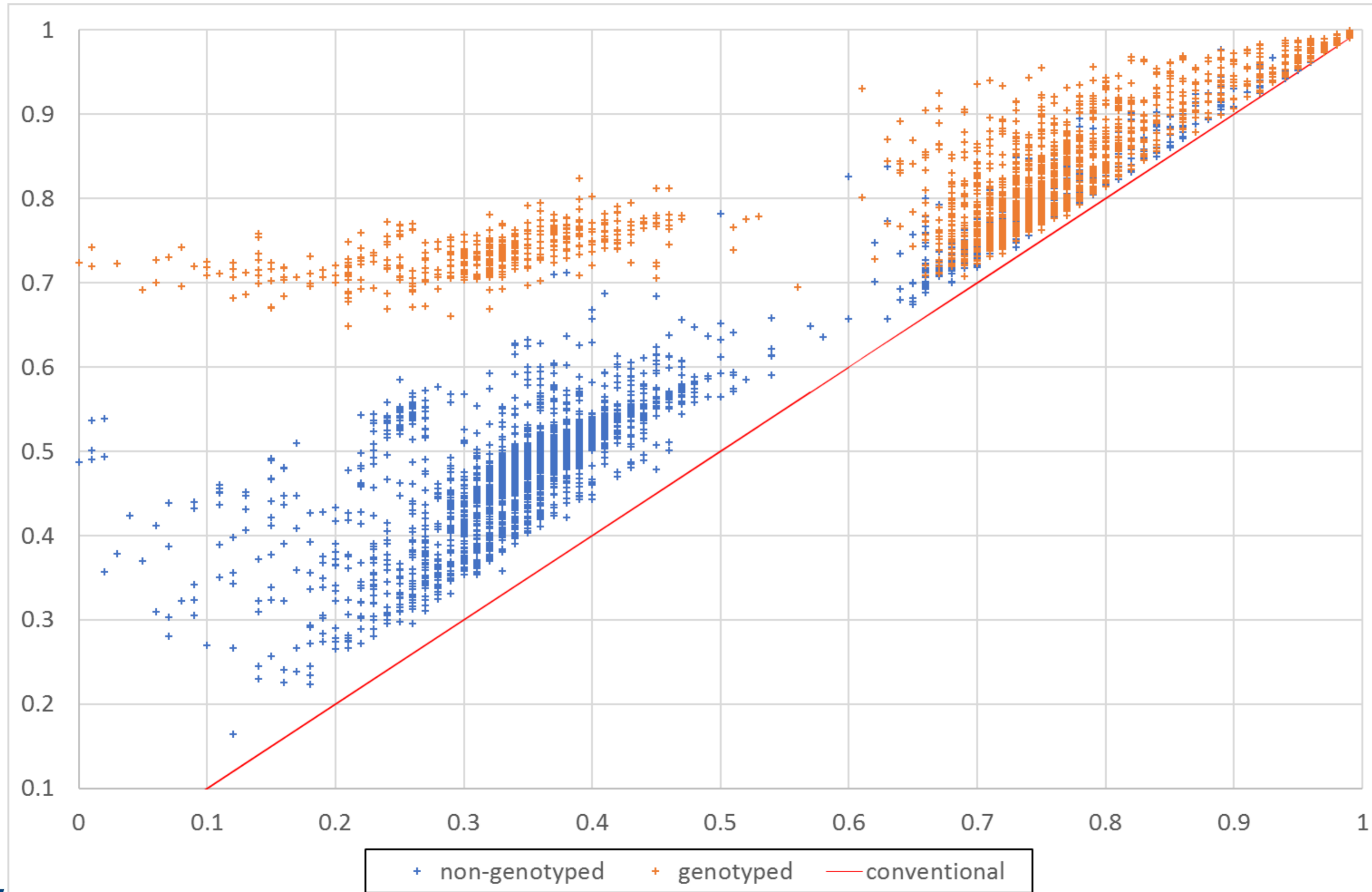 - Number of genotypes   : 115,378
  - Number of SNP   :   37,995

BETTER COWS | BETTER LIFE
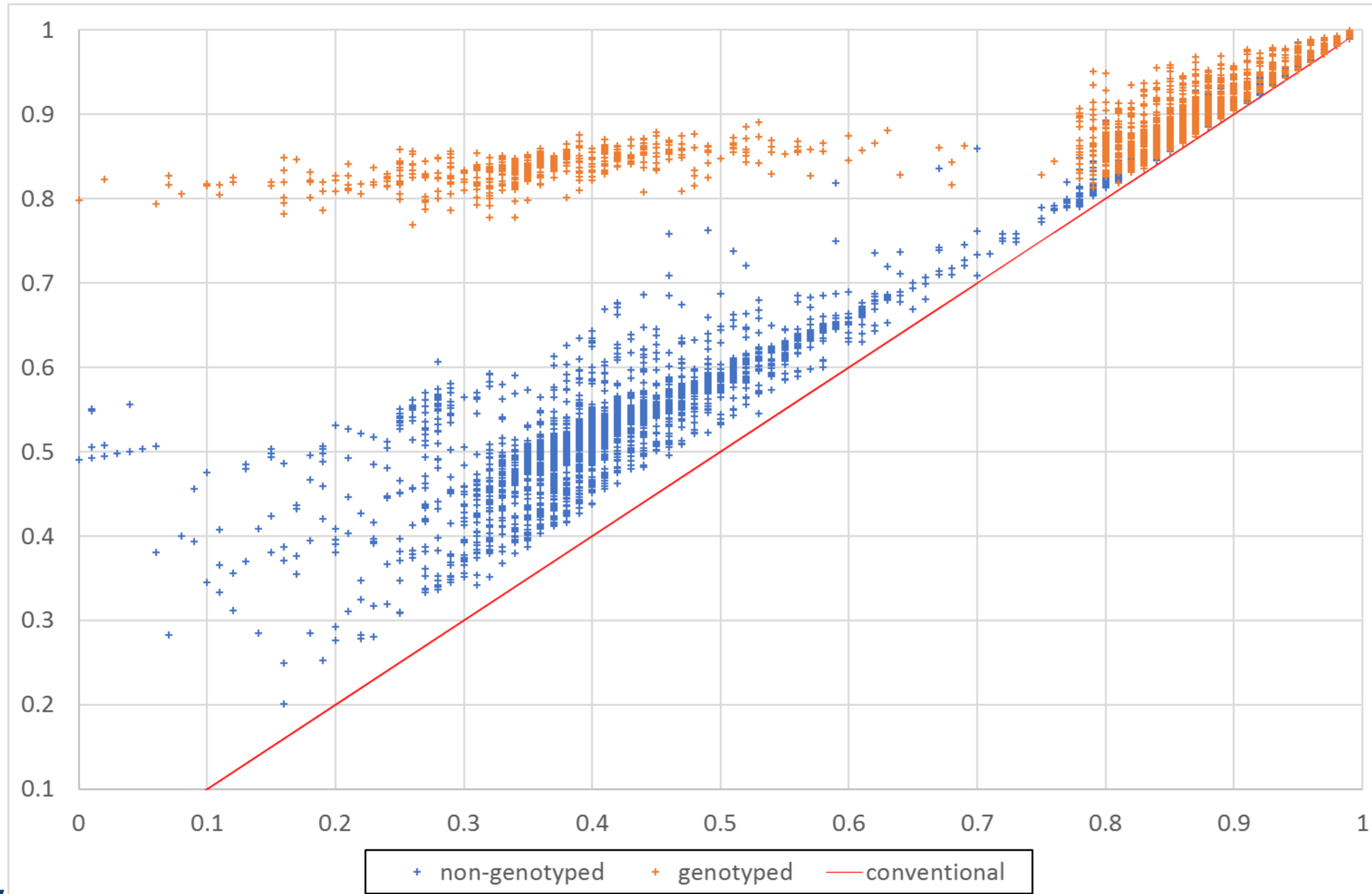
RESULTS: BULLS GEBV VS EBV MILKING SPEED

RESULTS: BULLS GEBV VS EBV MILKING SPEED

Young genotyped bulls (no daughters)
- Mean GEBV reliability : 0.81
- Mean added EDC : 71.0

+ non-genotyped   + genotyped   — conventional

CRV

BETTER COWS | BETTER LIFE
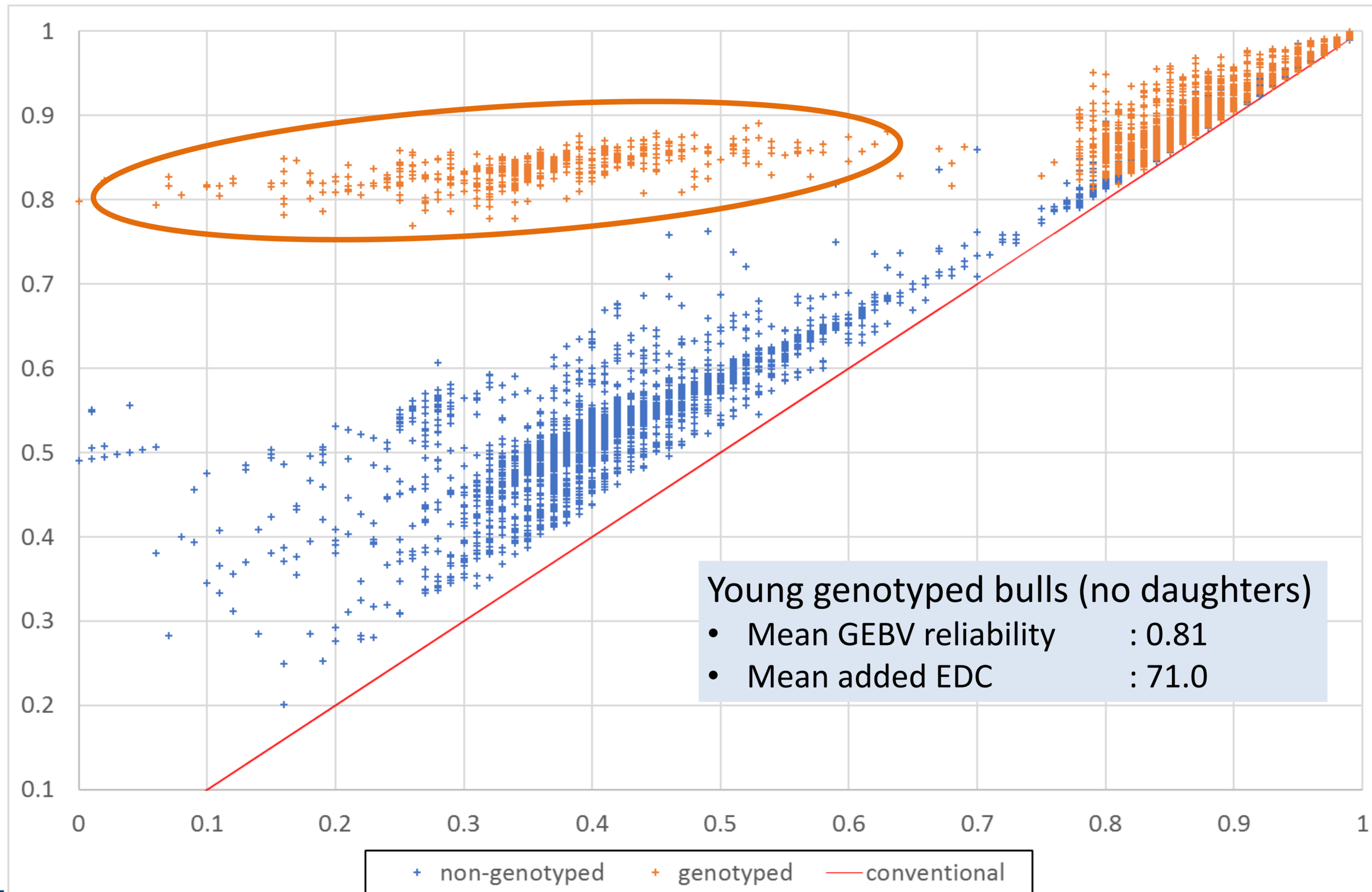
RESULTS: BULLS GEBV VS EBV TEMPERAMENT

Young genotyped bulls (no daughters)
- Mean GEBV reliability   : 0.72
- Mean added EDC   : 70.9
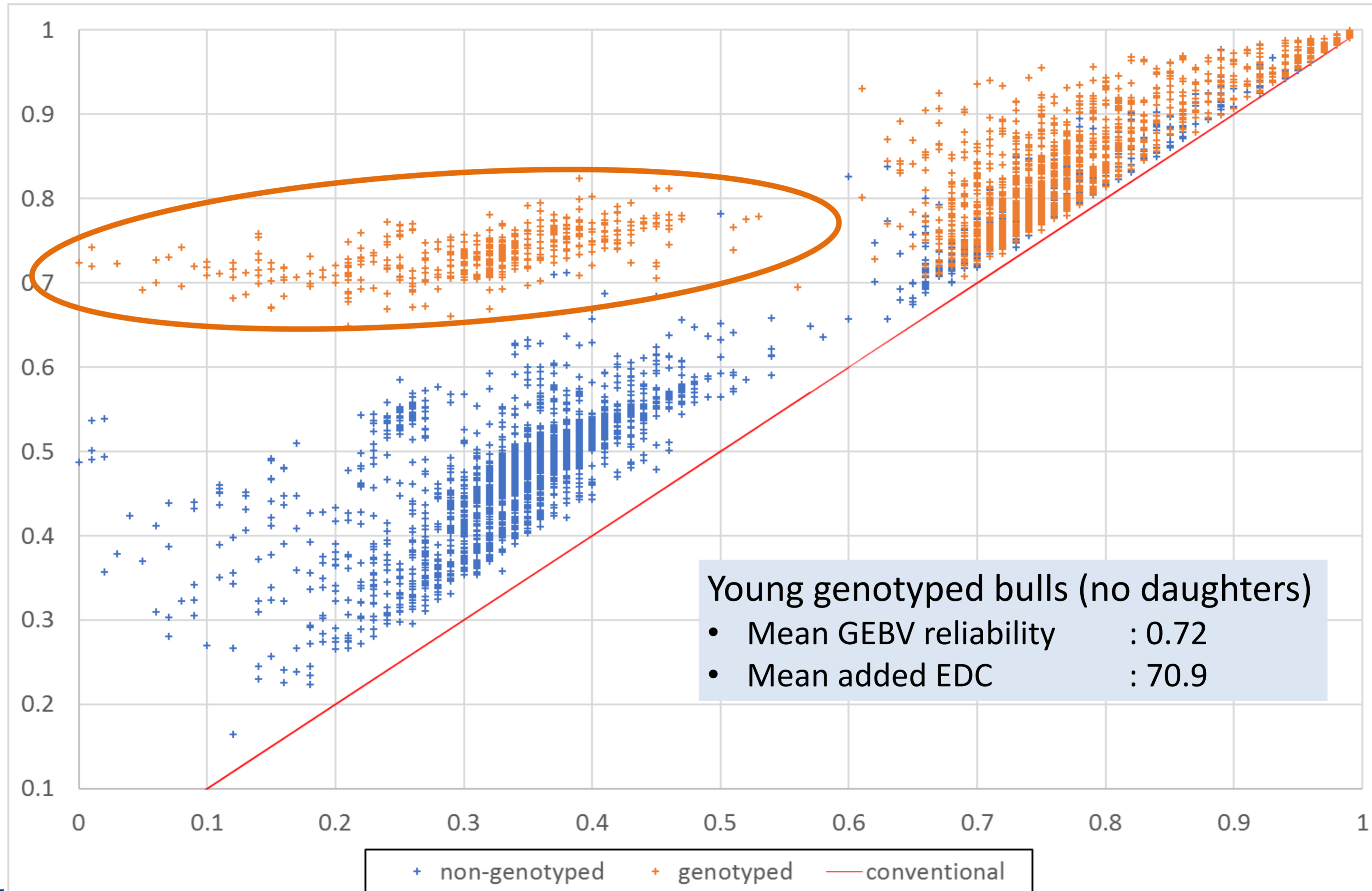
TEST DATA USED: TEMPERAMENT AND MILKING SPEED

- Overview traits:

Trait	EDC$_{add}$ (val 2018)		Mean rel. (GEBV YB aug '18)	
Milking speed	24.1		0.71	
Temperament	5.1		0.38	

BETTER COWS | BETTER LIFE

TEST DATA USED: TEMPERAMENT AND MILKING SPEED

- Overview traits:

Trait	EDC_{add} (val 2018)	EDC_{add} (ITB YB)	Mean rel. (GEBV YB aug '18)	Mean rel. (ITB YB)
Milking speed	24.1	71.0	0.71	0.81
Temperament	5.1	70.9	0.38	0.72

- Conclusion: Correction/scaling seems unavoidable

CRV

BETTER COWS | BETTER LIFE

DISCUSSION

DISCUSSION: RUNTIME

- Getting SNP reliabilities was most costly in terms of run time

  - ~ 55 minutes per trait

- Most evaluations are multiple trait

  - Example: Fertility evaluation has 40 traits   ~ 35h (at 120,000 genotypes of 37,995 SNP)
  - Conventional rels fertility   ~ 20h
  - Solution: run traits in parallel

- Applying protocol to obtain GEBV rels requires little time

  - Includes propagation in 6.2 mln non-genotype animals

- Results suggest the protocol is fit for use in routine evaluation.

CRV

DISCUSSION: RESULTS

- Propagation results in increased Grel vs conventional reliability
  - VanRaden&Wiggans algorithm gives satisfactory results

- Grel seems to overestimate reliability of GEBV
  - Possible cause 1: No deregression (yet) of conventional EDC
  - Possible cause 2: No scaling with 'realized' reliability

- Deregression not expected to reduce overestimation much

- Conclusion: Scaling of REL_{SNP} or EDC_{SNP} is necessary
  - Most important factor in estimation procedure
  - Possibly use 'added EDC' statistic from genomic validation

BETTER COWS | BETTER LIFE

DISCUSSION: SCALING

- Genomic validation produces two main statistics:
  - R^2_{DGV} : Mean genomic reliability of validation bulls (no daughters)
  - R^2_{BLUP} : Mean conventional reliability of validation bulls

- Additional statistic produced: mean added EDC
  - $\Delta_{EDC} = EDC(R^2_{DGV}) - EDC(R^2_{BLUP})$
  - Estimate of EDC's added to information by single genotype of bull without genotyped relatives

- Alternative scaling (will be tested):
  - $REL_{SNP} => EDC_{SNP}$
  - Get EDC_{SNP} for bulls in validation
  - $f = \Delta_{EDC} / mean(EDC_{SNP, val})$
  - $EDC_{dgv,i} = EDC_{snp,i} \times f \times (1 - r_{imp})$
    - Residual additive genetic variance already accounted for.

CRV

BETTER COWS | BETTER LIFE

DISCUSSION: CONCERNS

- Interbull protocol relies on a number of 'outside' parameters
  - Proportion of residual additive genetic variance k
  - Ratio of realized versus estimated reliability f
- Possibly these can be derived from validation results
- But: many more traits in routine evaluation than in genomic evaluation
  - E.g. lactation specific traits, predictor traits
- Values for k and f not readily available for all
  - Especially f seems pivotal to correct estimation of grel.
- Ideal: An approach that estimates correct REL_{SNP} or EDC_{SNP}
  - Would make possible a self-contained reliability estimation

CRV

BETTER COWS | BETTER LIFE

FINAL REMARKS

- Reliability calculations using Interbull protocol seems feasible for routine use.

- Given parameters used we get reasonable estimates

- Correct estimates are contingent on correct f value