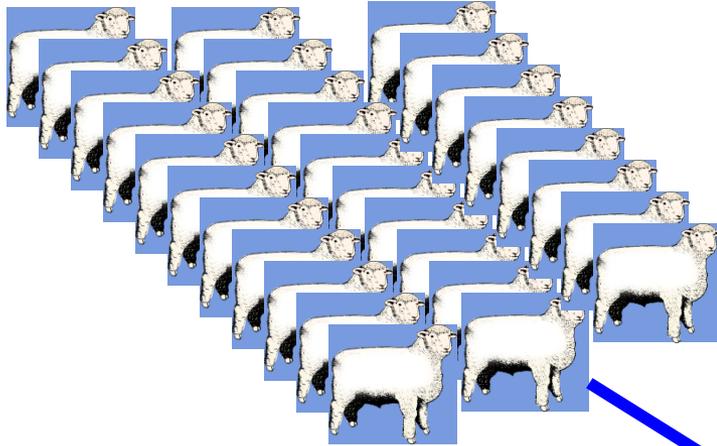
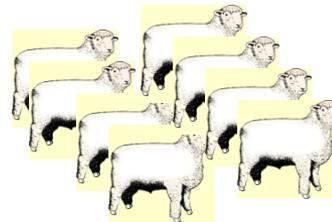
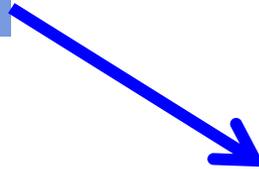


Genomic Prediction: basic idea



Reference population
measured and DNA tested



Young sires
Only DNA tested

To predict a trait EBV at a young age,

good for:

late traits

hard to measure traits

The questions

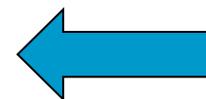
- How many records are needed in the reference population to achieve a certain accuracy?

But also:

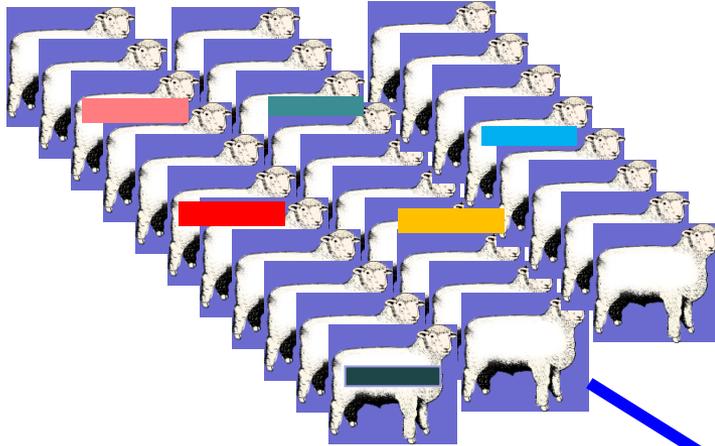
- What if you are more related to the reference?
- property of the reference population (heterogeneous, multi breed?)
- the value of closer relatives

Genomic prediction accuracy

- Derive from the model, e.g. PEV from GBLUP mixed model equations
- Validate with other EBVs or phenotypes
 - Validation population
 - Cross-validation
- Predict in advance based on theory and assumptions about population

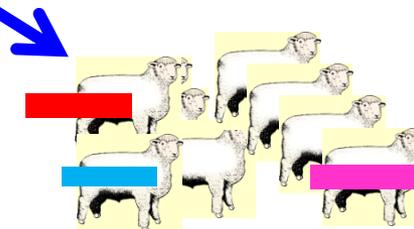


Genomic Prediction: basic idea



1) Reference population

diversity and length of segments → accuracy



2) young sire

Accuracy depends on:

Linkage Disequilibrium

Sharing haplotypes

Genomic Relationships

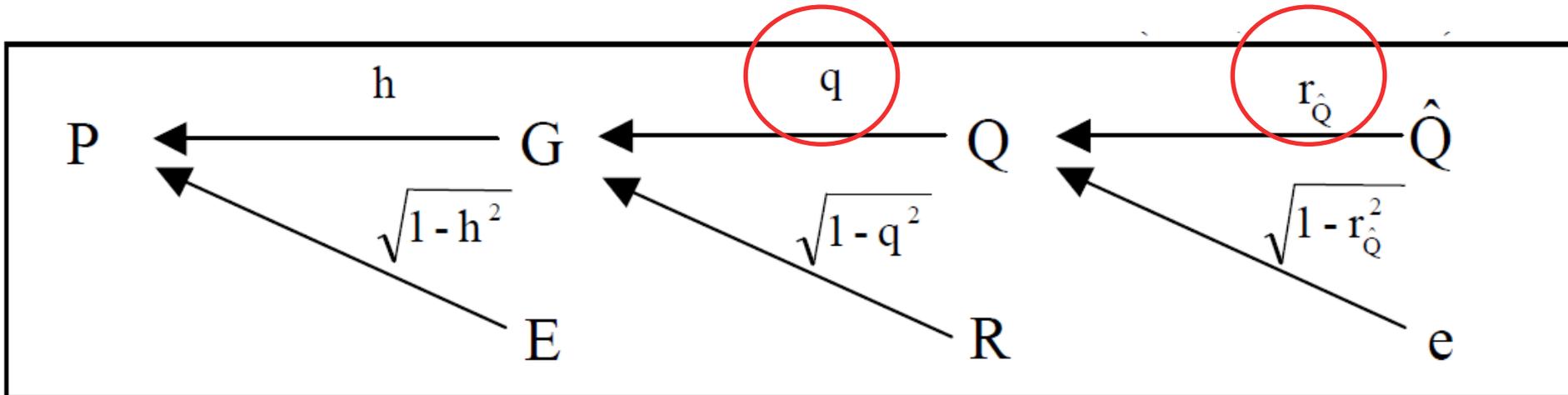
Genomic prediction accuracy *Using Goddard et al, 2011*

Depends on

i) Proportion of genetic variance at QTL captured by markers

i) Reliability of estimating marker effects

See also Dekkers 2007 (Path coefficient method)



Trait heritability = h^2

G = total BV

Q = genetic effects captured by marker(s)

R = residual polygenic effects

Model for phenotype: $P = G + E$

Model for BV: $G = Q + R$

Genomic prediction accuracy *Using Goddard et al, 2011*

Depends on

i) Proportion of genetic variance at QTL captured by markers q^2

ii) Reliability of estimating marker effects

r^2_{Qhat}

$$\begin{aligned} \text{Accuracy} &= \sqrt{q^2 \cdot r^2_{Qhat}} \\ &= q \cdot r_{Qhat} \end{aligned}$$



Genomic prediction accuracy Using Goddard et al, 2011

Depends on

- i) Proportion of genetic variance at QTL captured by markers $q^2 = M/(M_e + M)$

↳ Depends on marker-QTL LD

↳ Depends on

$M = \#$ markers

$M_e =$ 'effective number of chromosome segments'

$M_e = 2N_e Lk / \ln(2N_e)$

or is it...?

- i) Reliability of estimating marker effects

$$r^2_{Qhat} = \frac{V_{qhat}}{V_q} = \frac{N}{N + \lambda}$$

$$\lambda = M_e / q^2 \cdot h^2$$

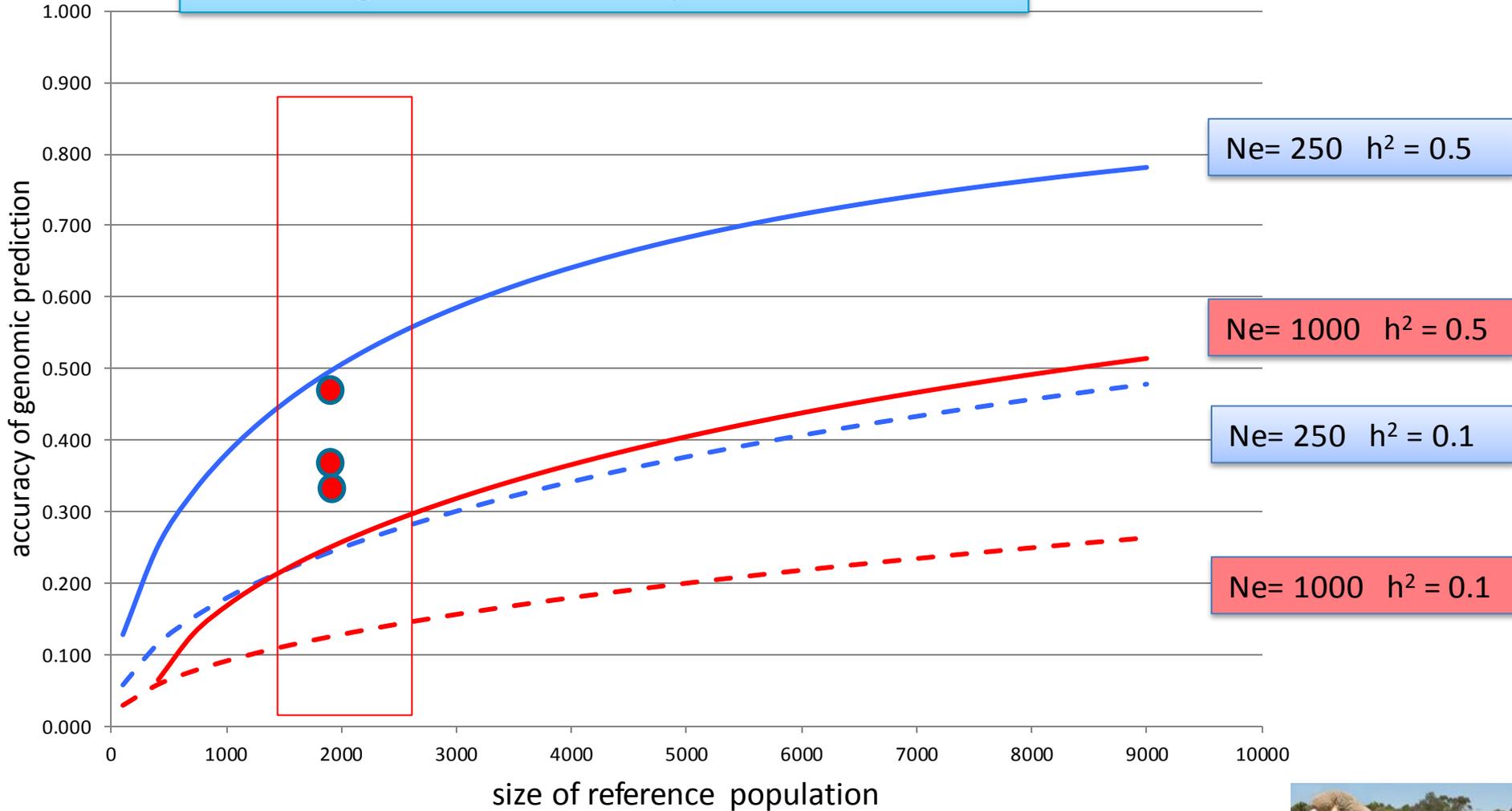
$$\text{Accuracy} = \sqrt{q^2 \cdot r^2_{Qhat}}$$

$$= q \cdot r_{Qhat}$$



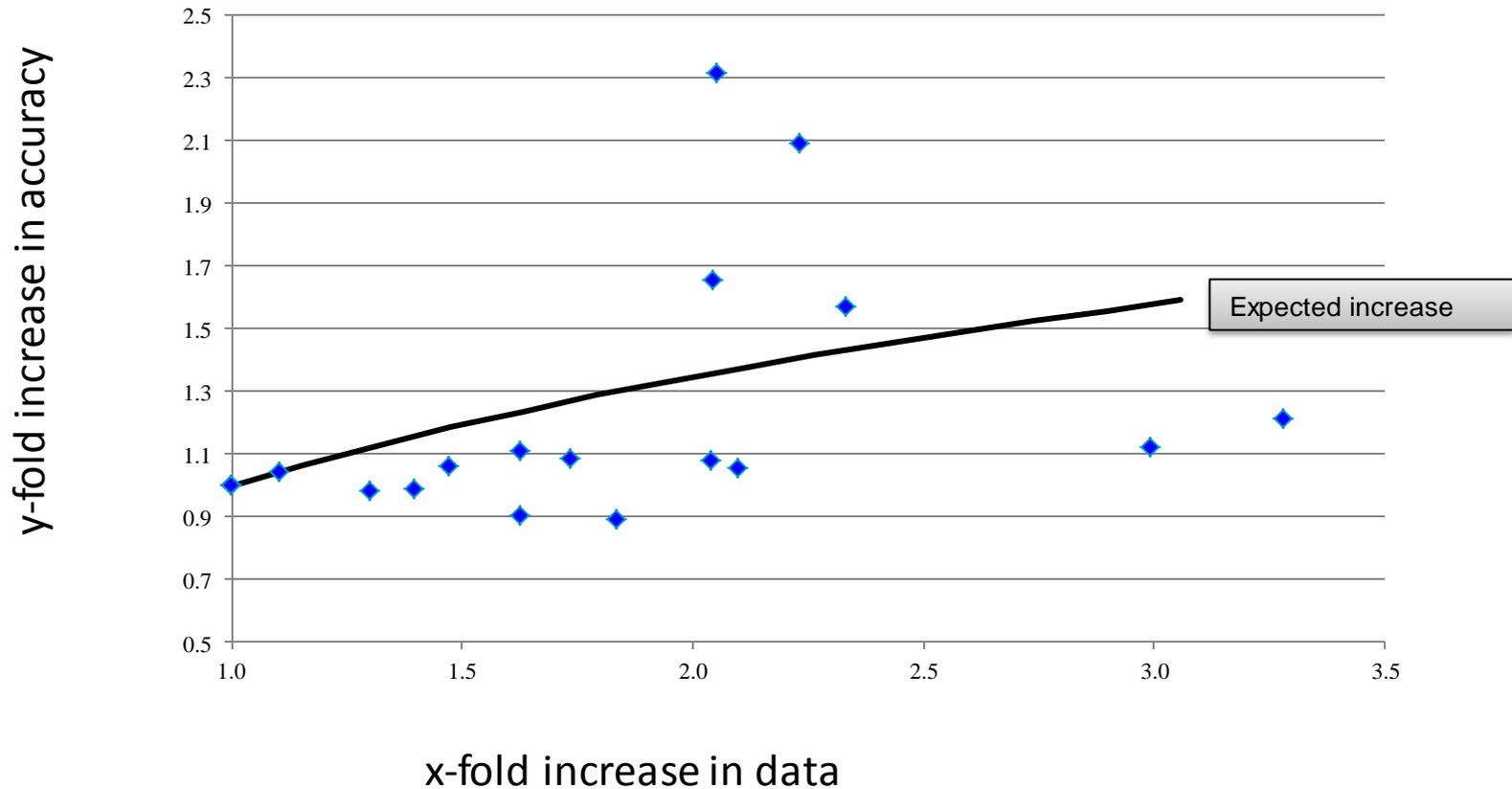
Genomic prediction accuracy *Using Goddard et al, 2011*

Did we get what we expected?



Validating 'Genomic Prediction Accuracy'

More data: does accuracy increase as expected?



Relationship with reference population

Clark et al 2011

Method	Close	Distant	Unrelated
	Ped 0 - 0.25 Genom 0.08 – 0.35	0 - 0.125 0.08 – 0.26	0 - 0.05 0.08 – 0.16
BLUP- Shallow pedigree	0.39	0.00	0.00
BLUP- Deep Pedigree	0.42	0.21	0.04
gBLUP	0.57	0.41	0.34

Additional accuracy from family info

'baseline accuracy': graphs predict 0.36
for $N_e=100$, $N=1750$, $h^2=0.3$

Selection Index principles

Single Trait selection index calculation using genomic testing

Parameters	
Heritability	0.3
Repeatability of subsequent records	0.3
c-squared (among full sibs)	0.15
Reliability of genomic test (prop expl)	0.09091

Run

Information used	Nr.Records	Genotyped	Index Wght		Value of Variate	
			Pheno	Geno	Pheno	Geno
nr of own records	1	1	0.214	0.550	14.7%	2.7%
nr. of records on dam	1	0	0.082	-	2.1%	-
nr of records on sire	1	0	0.030	-	0.2%	-
nr of fulls sib records	4	0	0.186	-	3.3%	-
nr. of half sib records (excl. full sibs)	40	0	0.347	-	3.0%	-
nr. of progeny	0	0	-	-	-	-

Accurcay of EBV **0.707**

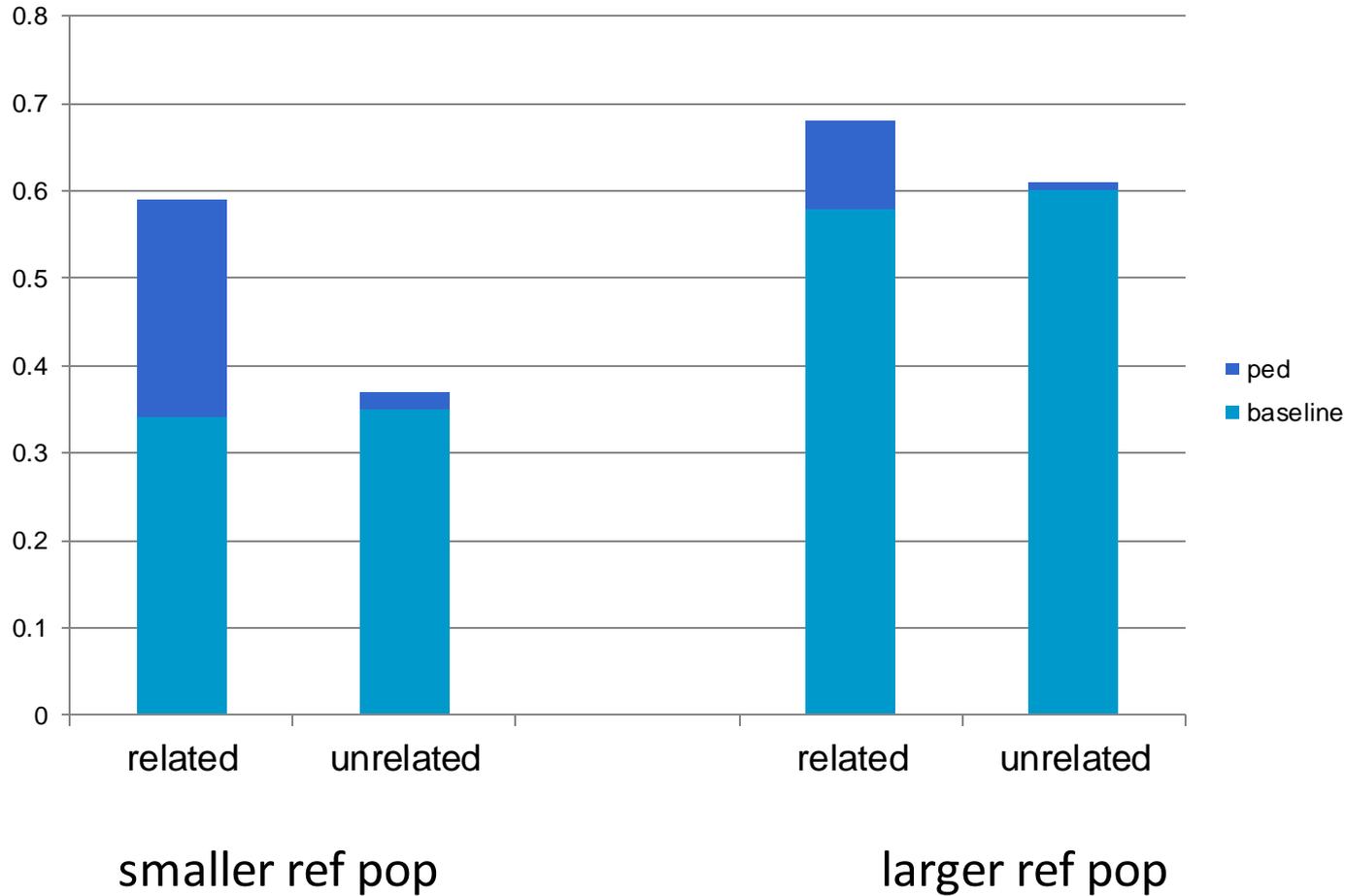
Accuracy of Pheno EBV **0.689**

Accuracy of MBV **0.302**

correlation EBV FS **0.797**

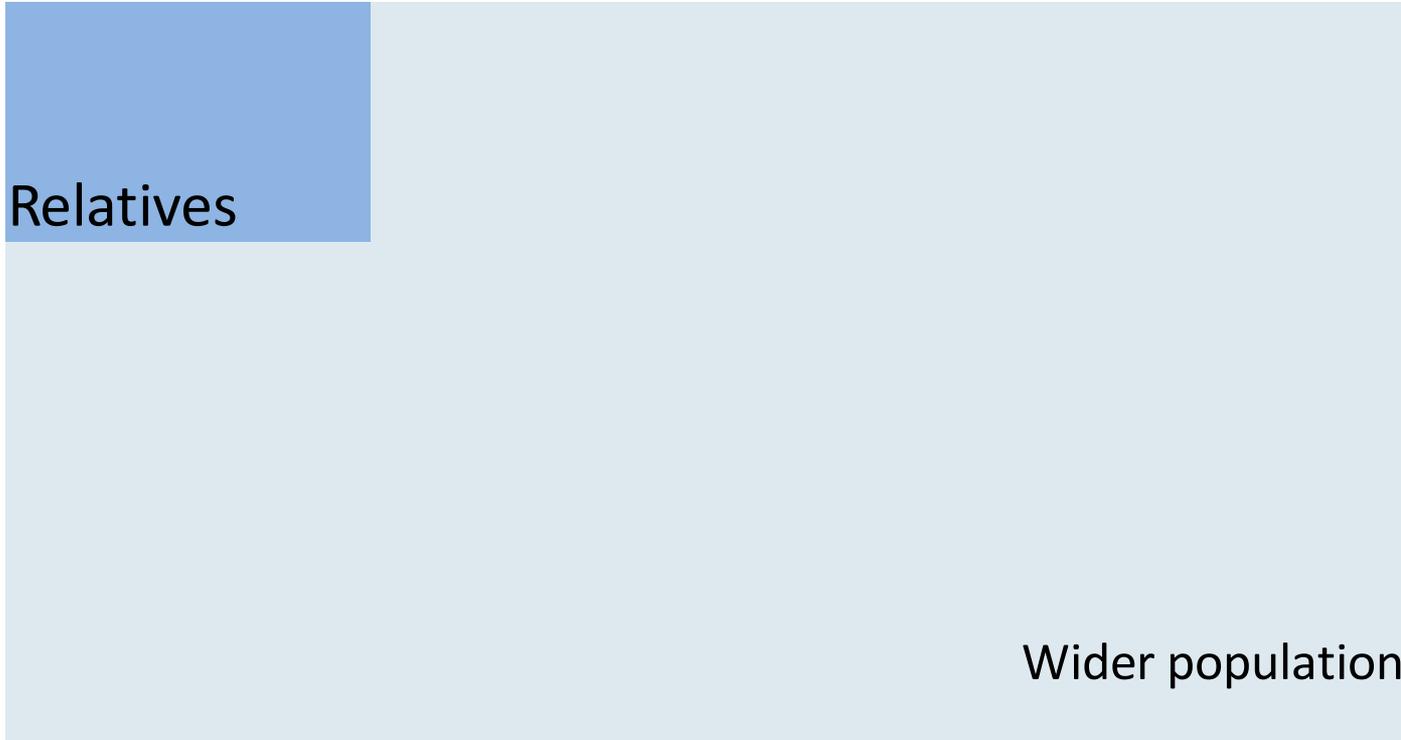
correlation EBV HS **0.440**

Relatedness matters more if the reference population is smaller



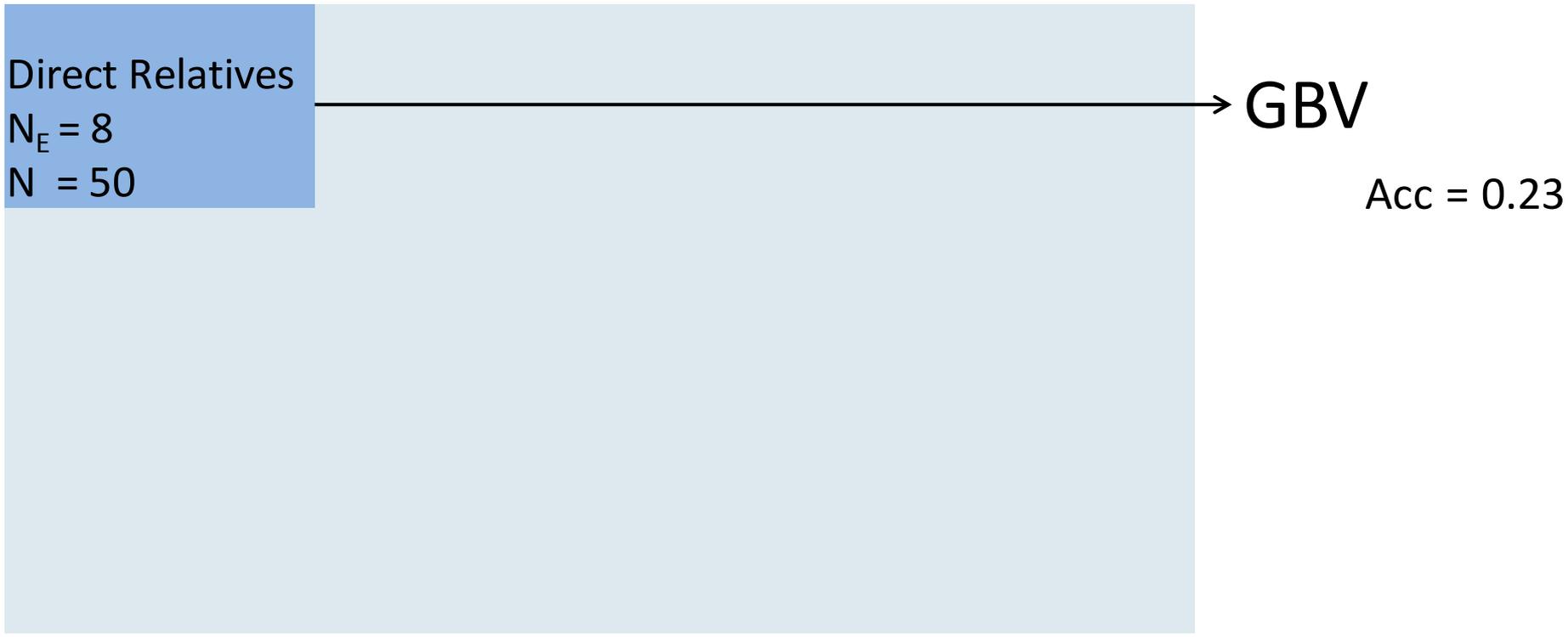
(hypothesis)

A reference population may have relatives

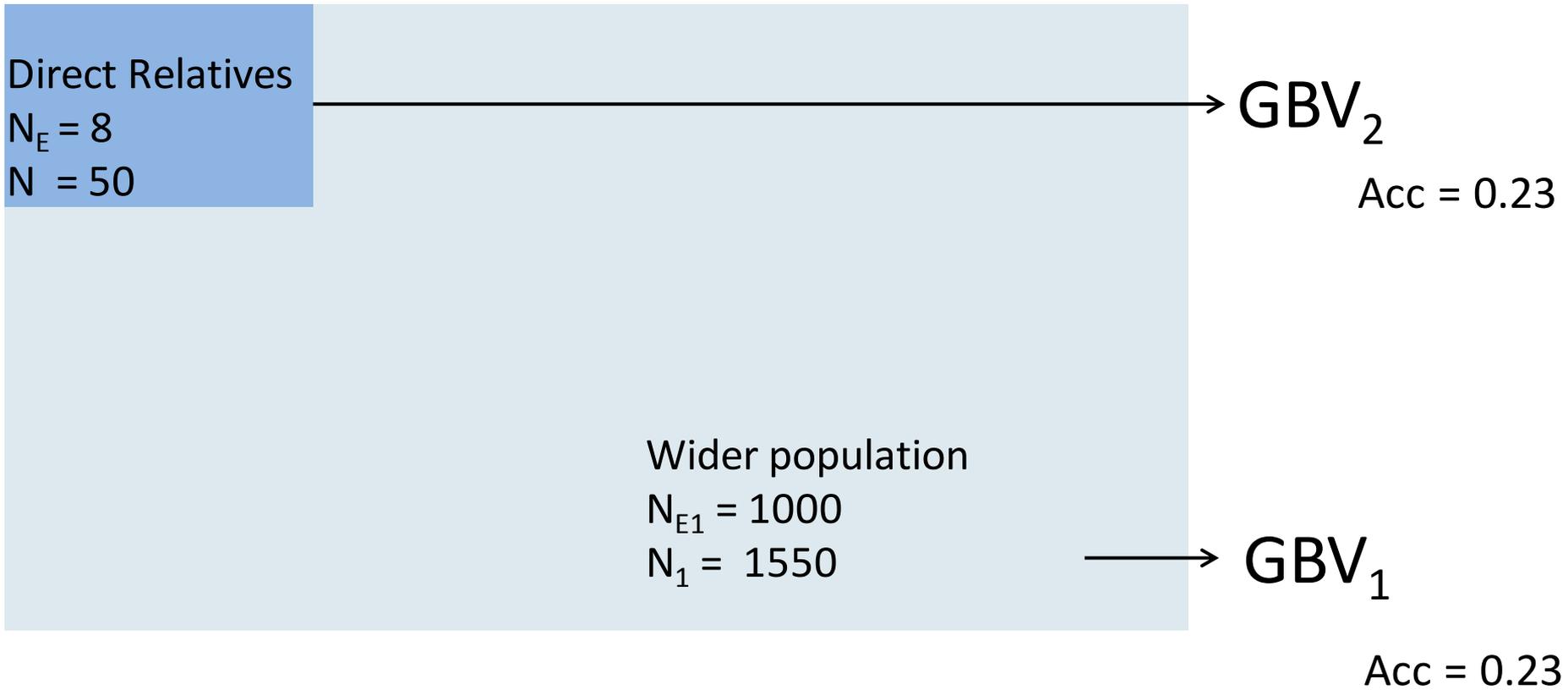


'Relatedness' can be represented by effective size

Hayes et al 2009



Information from different subsets can be combined

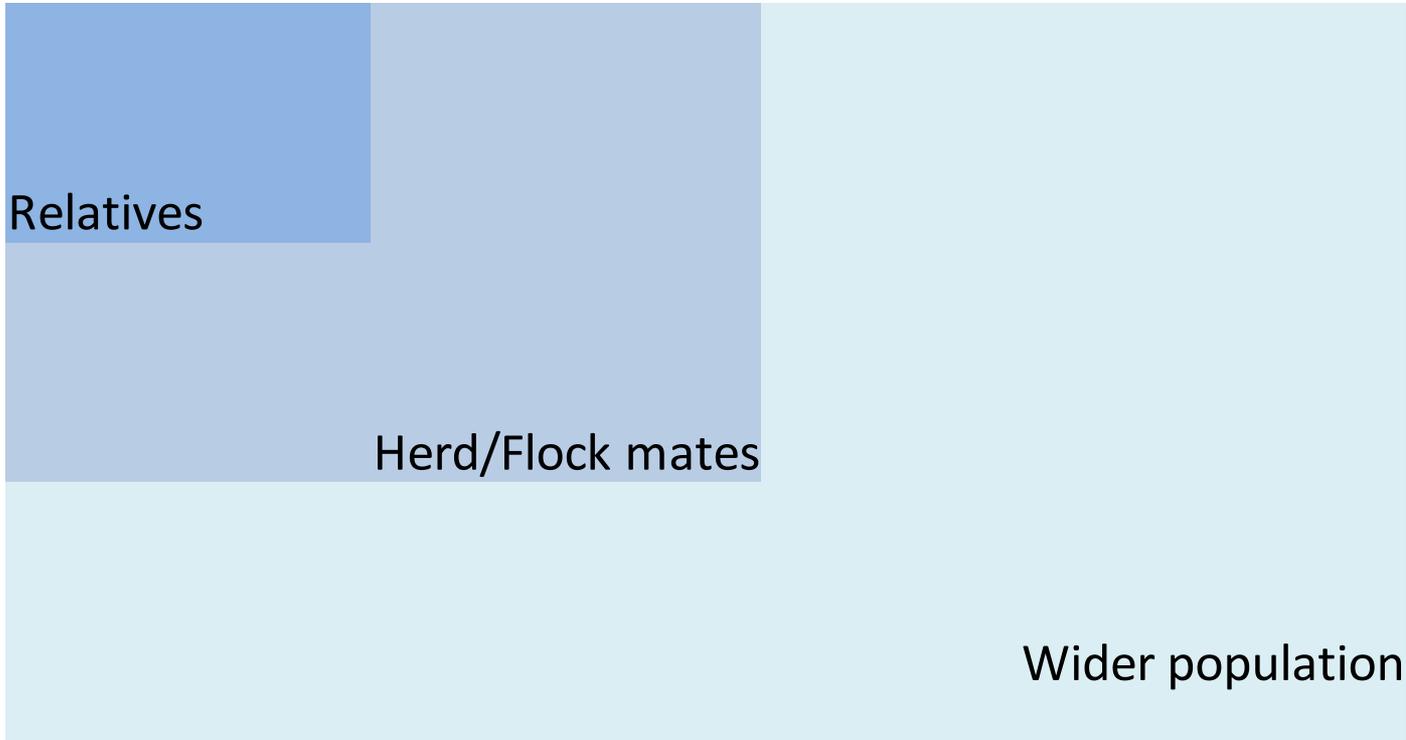


Calculate overall accuracy using selection index

$$GBV = \sum b_i GBV_i \quad Acc = 0.31$$

Using a stratified reference population

-populations are not homogeneous



Using a stratified reference population

-populations are not homogeneous

Direct Relatives

$$N_{E3} = 8$$

$$N_3 = 50$$

Own Herd

$$N_{E2} = 50$$

$$N_2 = 400$$

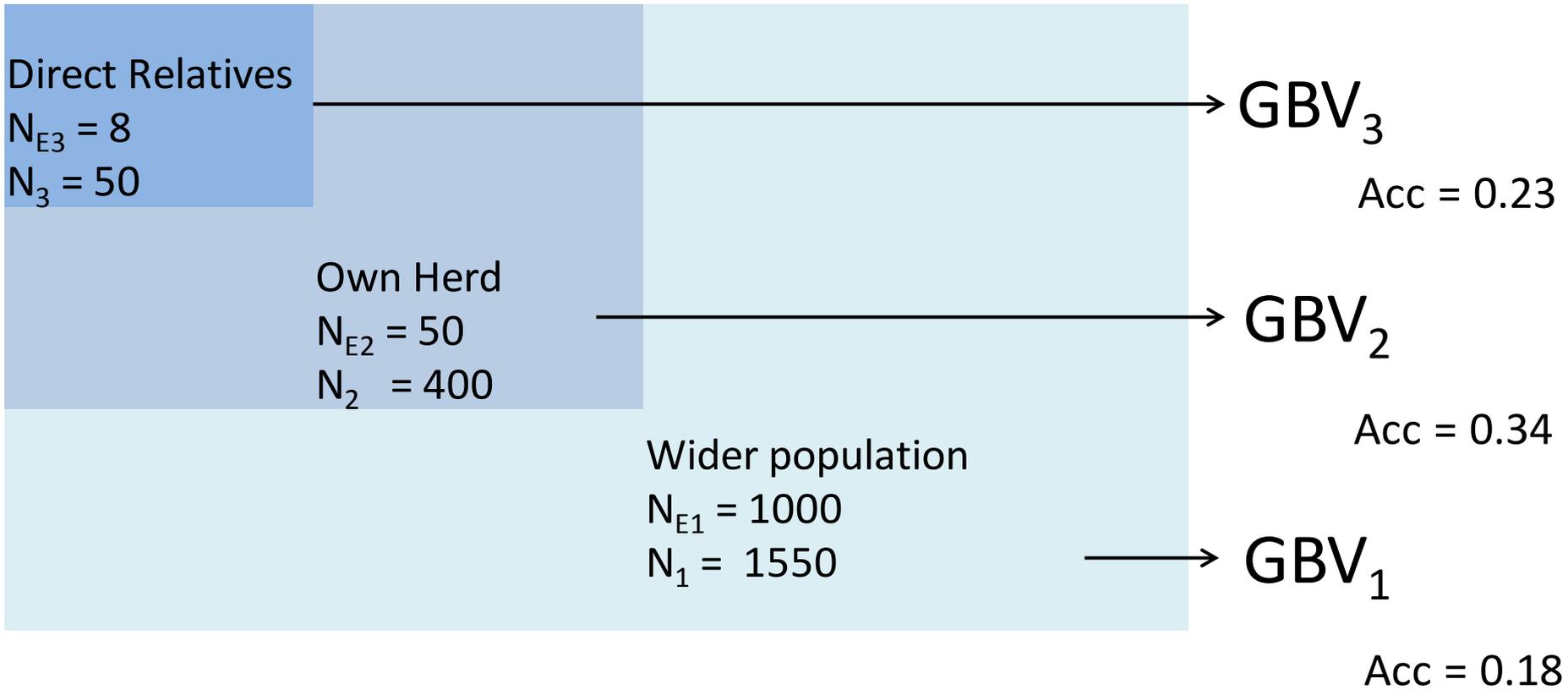
Wider population

$$N_{E1} = 1000$$

$$N_1 = 1550$$

Using a stratified reference population

-populations are not homogeneous

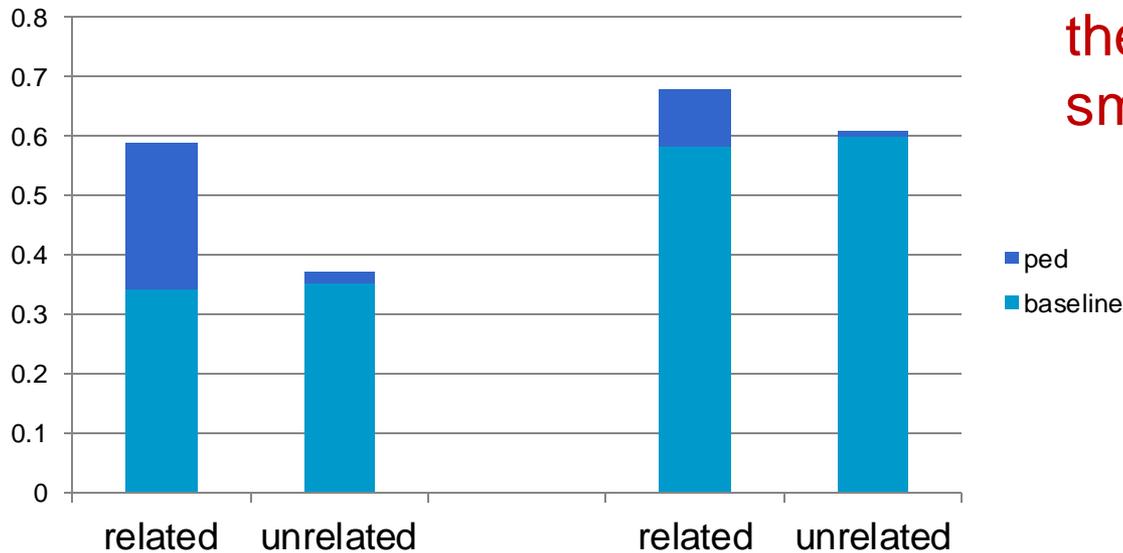


Calculate overall accuracy using selection index

$$GBV = \sum b_i GBV_i \quad Acc = 0.42$$

$NE_1 = 1000$

N_1	Value of information source			GBV accuracy		diff
	breed (N1)	flock (400)	relatives (50)	all info	breed only	
2,000	16%	52%	21%	0.43	0.22	95%
5,000	31%	39%	15%	0.47	0.32	48%
10,000	45%	26%	10%	0.53	0.42	26%



Relatedness matters more if the reference population is smaller

hypothesis confirmed

$$NE_1 = 1000$$

	Value of information source			GBV accuracy		
N_1	breed (N1)	flock	relatives	all info	breed only	diff
		400	50			
2,000	16%	52%	21%	0.43	0.22	95%
5,000	31%	39%	15%	0.47	0.32	48%
10,000	45%	26%	10%	0.53	0.42	26%
N_1	breed (N1)	flock	relatives	all info	breed only	diff
		100	10			
2,000	48%	36%	48%	0.28	0.21	36%
5,000	68%	19%	68%	0.36	0.31	15%
10,000	79%	11%	79%	0.45	0.41	7%

With fewer relatives the reliance on the reference population increases

$NE_1 = 1000$

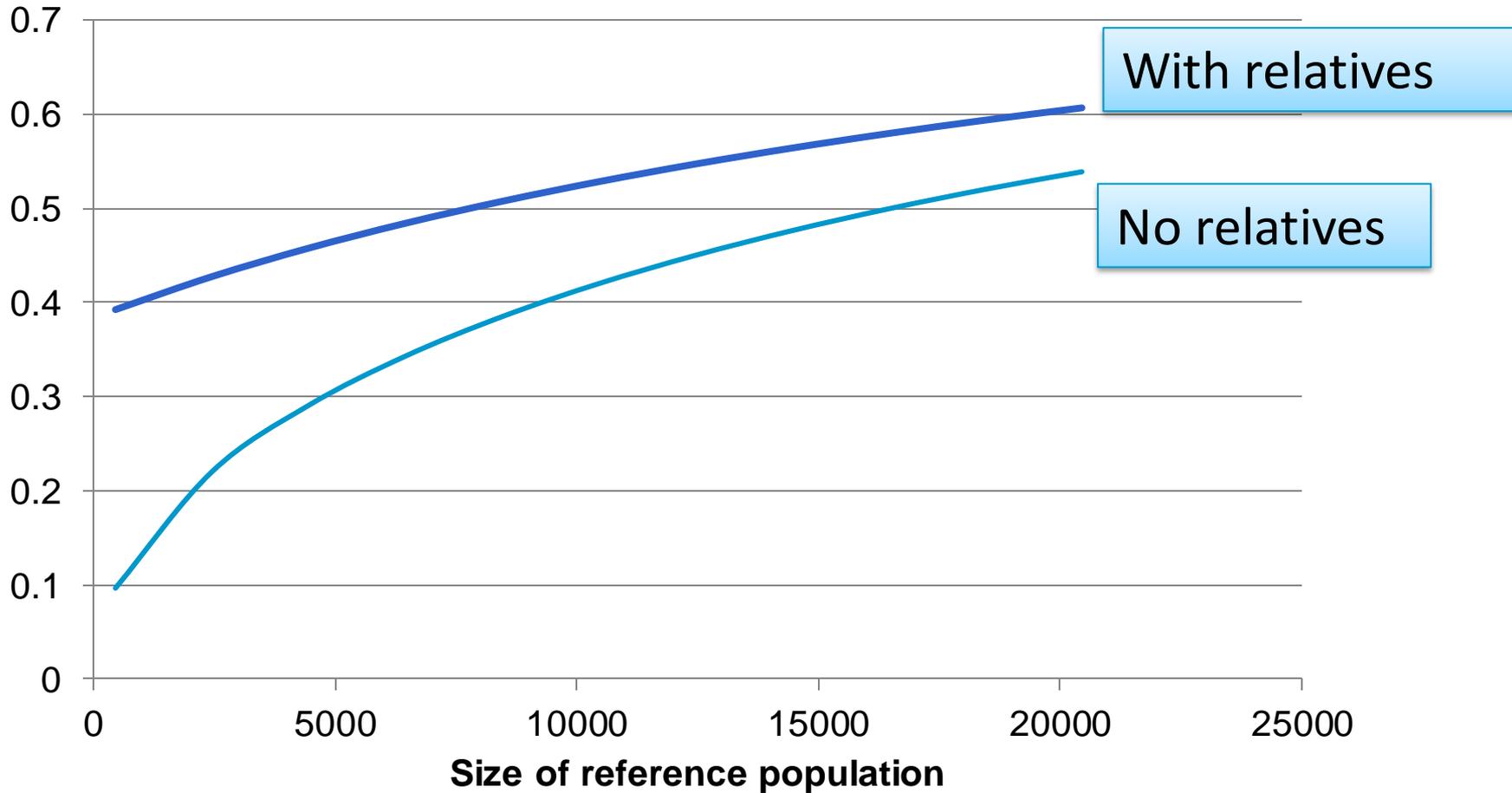
	Value of information source			GBV accuracy		
N_1	breed (N1)	flock (400)	relatives (50)	all info	breed only	diff
2,000	16%	52%	21%	0.43	0.22	95%
5,000	31%	39%	15%	0.47	0.32	48%
10,000	45%	26%	10%	0.53	0.42	26%

$NE_1 = 200$

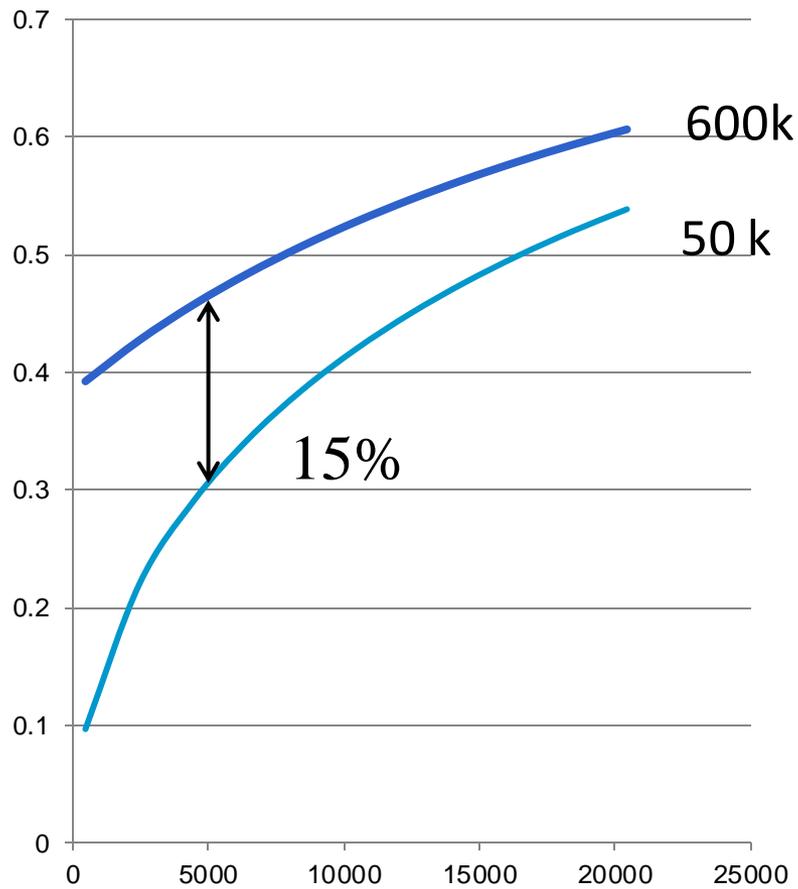
N_1	breed (N1)	flock (400)	relatives (50)	all info	breed only	diff
2,000	45%	26%	10%	0.53	0.45	18%
5,000	62%	12%	5%	0.64	0.60	7%
10,000	72%	5%	2%	0.74	0.72	3%

With less diverse populations the relatives matter a lot less

The effect of a larger reference population.

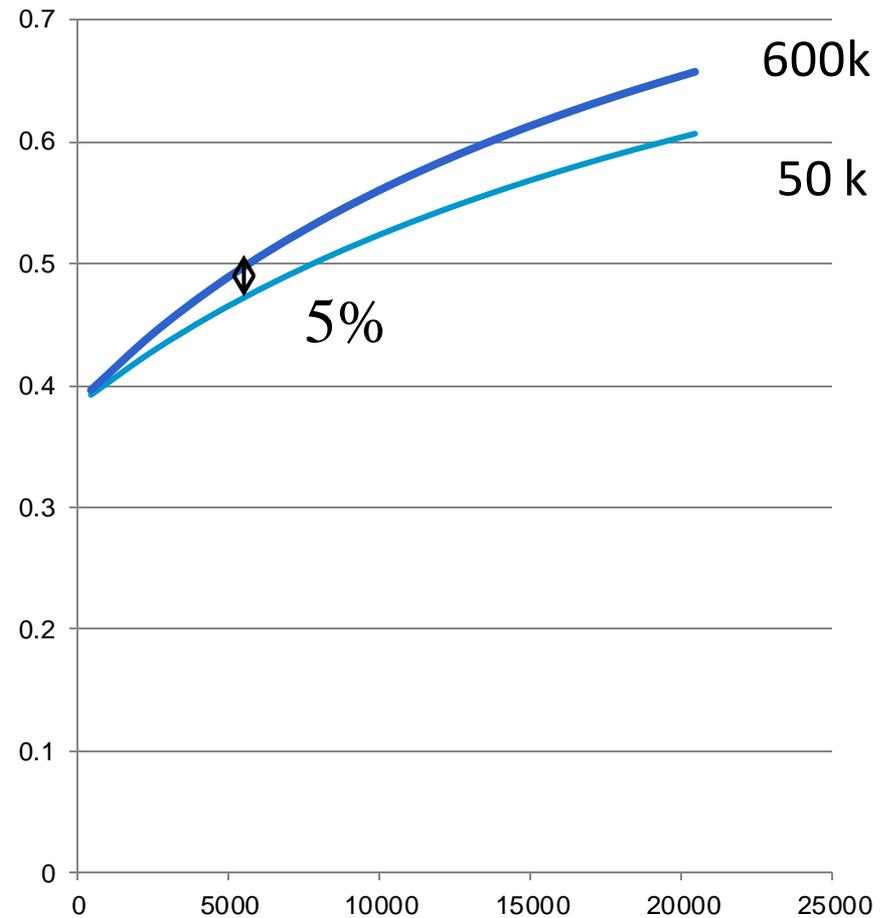


The effect of denser marker panels



Size of the reference population

No relatives



Size of the reference population

With relatives

Conclusions

- Theory exists to predict genomic prediction accuracy in advance: depends on population diversity, nr records
- Reference populations are heterogeneous, with closer as well as distant relatives
- Relatives will increase accuracy and decrease reliance on wider reference population (and denser marker) but that information has a shorter life

