Design of reference populations
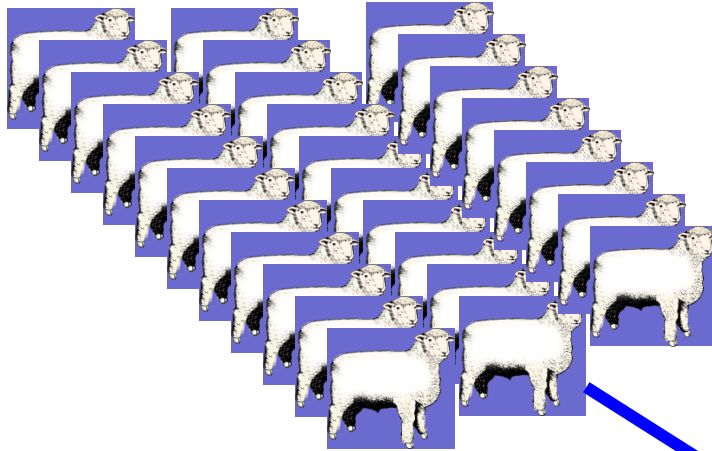
Julius van der Werf
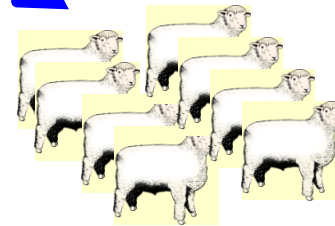
CRC for Sheep Industry Innovation
School of Environmental and Rural Science, UNE, Armidale, NSW

Genomic Prediction: basic idea

1) Somebody (else) measures lots of sheep, and their DNA
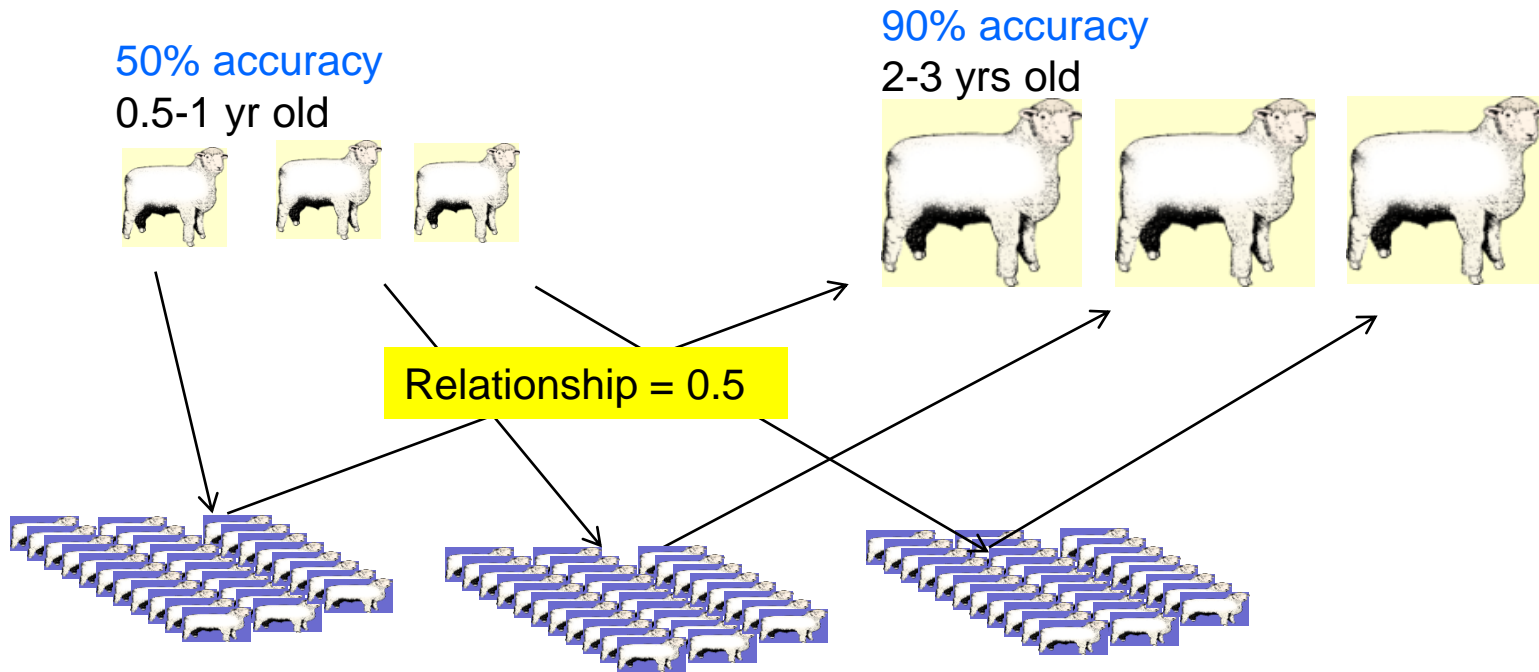→ Reference population

2) A breeder tests DNA on young rams

Prediction from DNA → genomic breeding values - GBV

GBV + Current ASBV → Improved ASBV
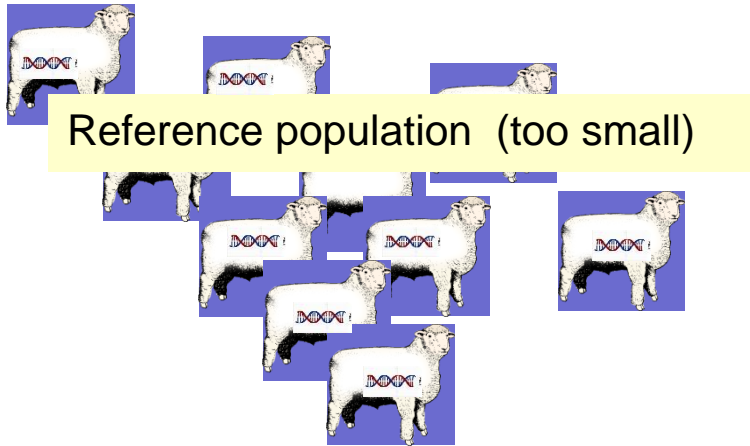
Merit depends on trait measurability

une

Compare: Progeny Testing

50% accuracy
0.5-1 yr old

90% accuracy
2-3 yrs old

Relationship = 0.5

Each progeny group only informs one sire

Genomic Testing

Reference population (too small)
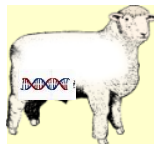
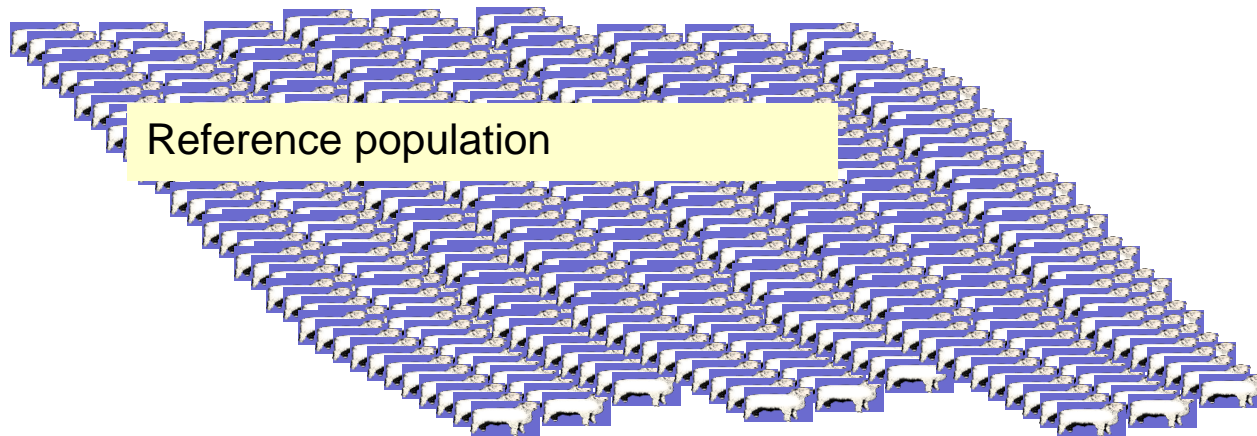Relationship = 0.02……0.5

use information on "relatives"
while sire is still young

51% accuracy
0.5-1 yrs old

une

Genomic Testing

Reference population

Relationship = 0.02……0.5

70% accuracy
0.5-1 yrs old

UNE
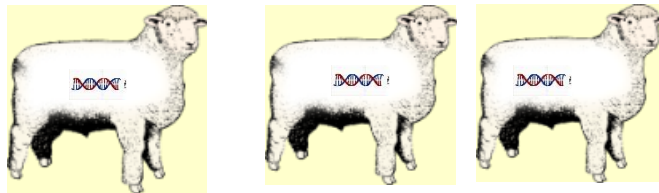
Benefits - Dairy

- Extra gain ~100%
- Breeding objective dominated by sex-limited trait
- No more progeny testing (save money)
- Very much shorter generation intervals
- More use of reproductive technologies
- Potential to select on hard to measure traits
  but only if these are being measured!
- Commercial males have more chance to be selected
- AI companies can easily afford testing
- Widely used in the industry

Benefits - Beef

- – Extra gain ~25-50%
- – Breeding objective has some hard to measure traits
- – More emphasis on carcass and meat, less on growth
- – More emphasis on females reproductive rate
- – Somewhat shorter generation intervals
- – More use of reproductive technologies
- – Potential to select on hard to measure traits
   but only if these are being measured!
- – Genotyping cost can be high for breeders

- – Who pay for the reference population?

Benefits – Pigs & Poultry

- – Extra gain ~50%?
- – More emphasis on meat quality, Feed Efficiency?
- – Sex limited traits
- – shorter generation intervals in layers
- – Potential to select on hard to measure traits but only if these are being measured!
- – Genotyping cost can be high?

Benefits - Sheep

- Extra gain ~25-50%
- Breeding objective has some hard to measure traits
- More emphasis on: carcass and meat, less on reproductive rate, 'lifetime wool', parasite resistance
- Somewhat shorter generation intervals
- Some more use of reproductive technologies
- Potential to select on hard to measure traits
  but only if these are being measured!
- Genotyping cost can be high for breeders
- Implemented in Australia, New Zealand

- Who pay for the reference population?

Setting up reference populations

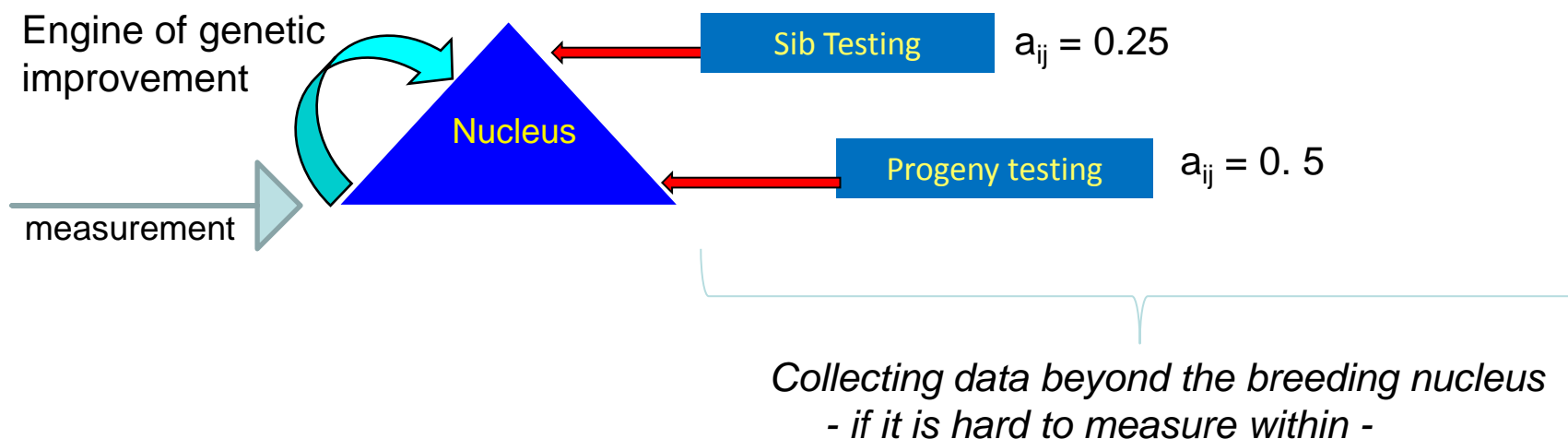| Trait is already measured	Early measurement	Late Measurement
YES	No Need	Use industry data (milk, fertility, late wool)
NO	Create Reference population (slaughter)	Create Reference population

Genomic selection has affected the need for phenotyping !

more…not less
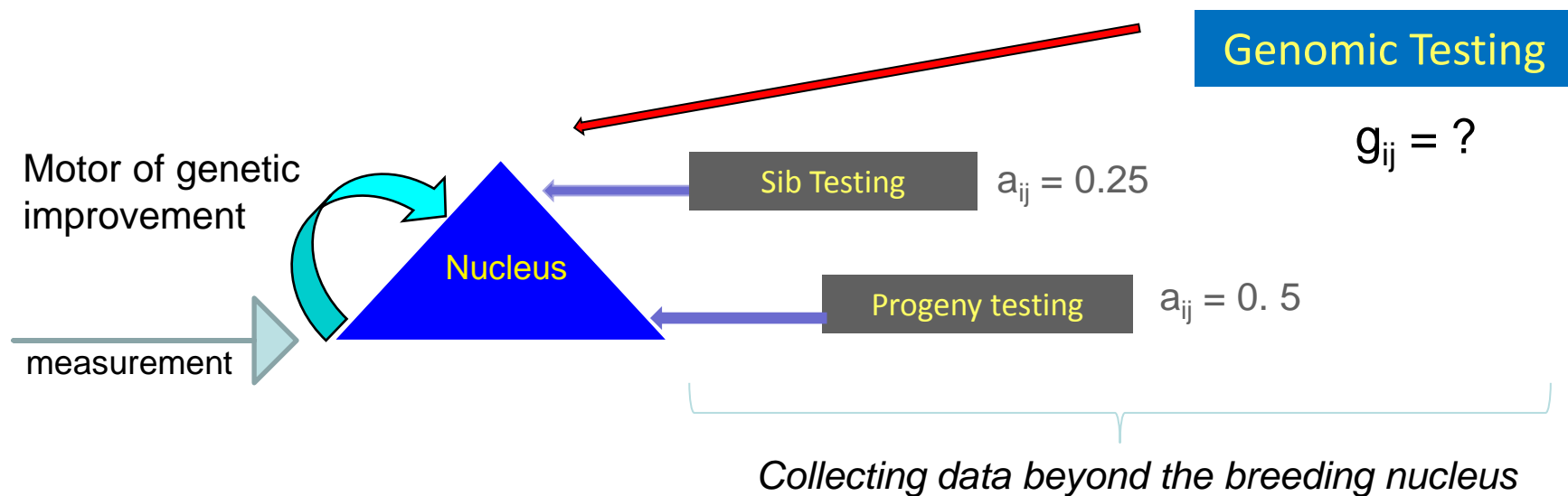
Who pays?

une

Design of a reference population

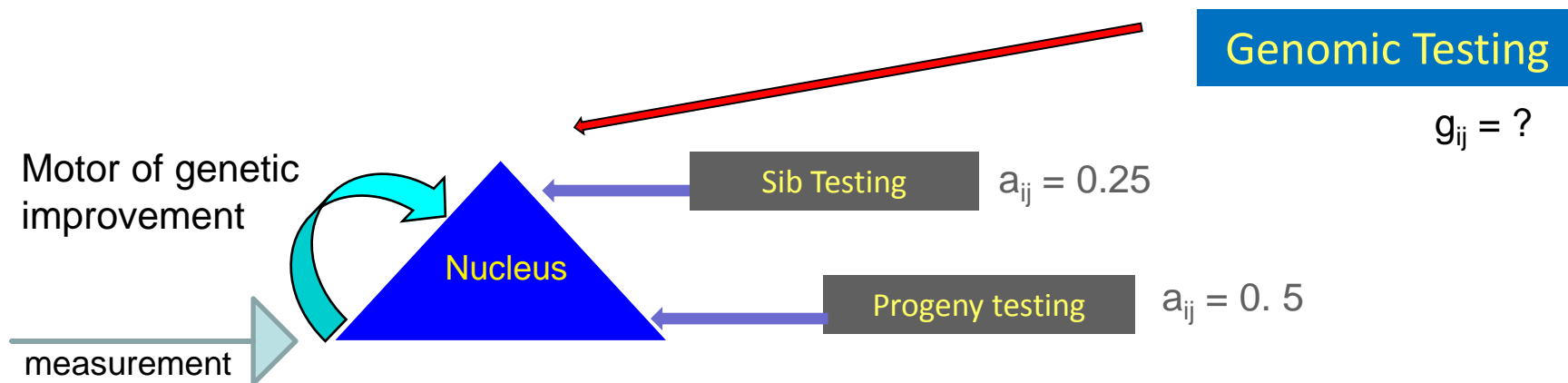# Investing in information for genetic improvement pre-genomics

Engine of genetic improvement

measurement

Nucleus

Sib Testing $a_{ij} = 0.25$

Progeny testing $a_{ij} = 0.5$

*Collecting data beyond the breeding nucleus
- if it is hard to measure within -*

Investing in information for genetic improvement

Genomic Testing

$g_{ij} = ?$

Motor of genetic improvement

Nucleus

Sib Testing $a_{ij} = 0.25$

Progeny testing $a_{ij} = 0.5$

measurement

Collecting data beyond the breeding nucleus

Investing in information for genetic improvement

Genomic Testing

$g_{ij} = ?$

Motor of genetic improvement

Nucleus

Sib Testing — $a_{ij} = 0.25$

Progeny testing — $a_{ij} = 0.5$

measurement

Measure outside nucleus if traits - can not be measured within
- carcass, eating quality, late wool

otherwise, reference population can be nucleus

Genomic selection has an advantage over sib or progeny test selection because

1. the information comes earlier
2. can afford to test more distant relatives

Design of Reference Population

Nucleus

Genomic Testing
Relationship paradigm

$a_{ij} > 0.05$

Genomic Testing
LD paradigm

$a_{ij} > 0.0000001$
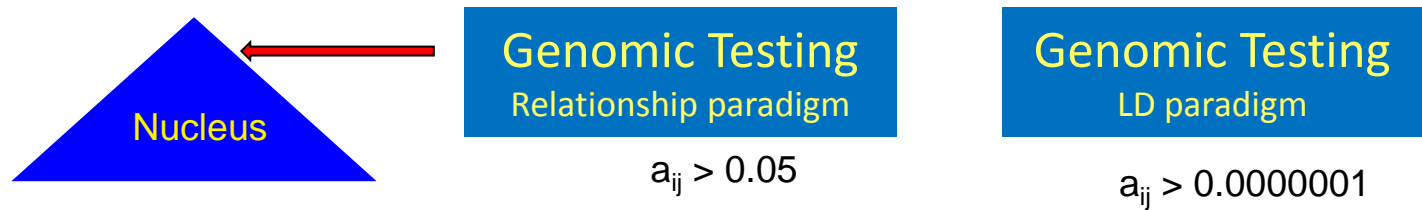
Relationship paradigm

Need relatives in reference
Need to keep reference 'up to date'
Denser markers maybe of limited benefit
Accuracy limited by relationships and # of relatives
Consider to use IBD inference

LD paradigm

May achieve prediction across breeds
Reference population of long lasting benefit
Accuracy limited by marker density and size of reference
Requires detectible average effects
across wide range of genetic background

Summarizing Genomic Prediction
- What information is used?

- Based on very many small – genomic- relationships

- Does not require 'direct relatives' to be tested

- Can be based on distant relatives 'some generations away'

- …..but the number of small relatives needs to be large (thousands)

- Can not predict across breed

une

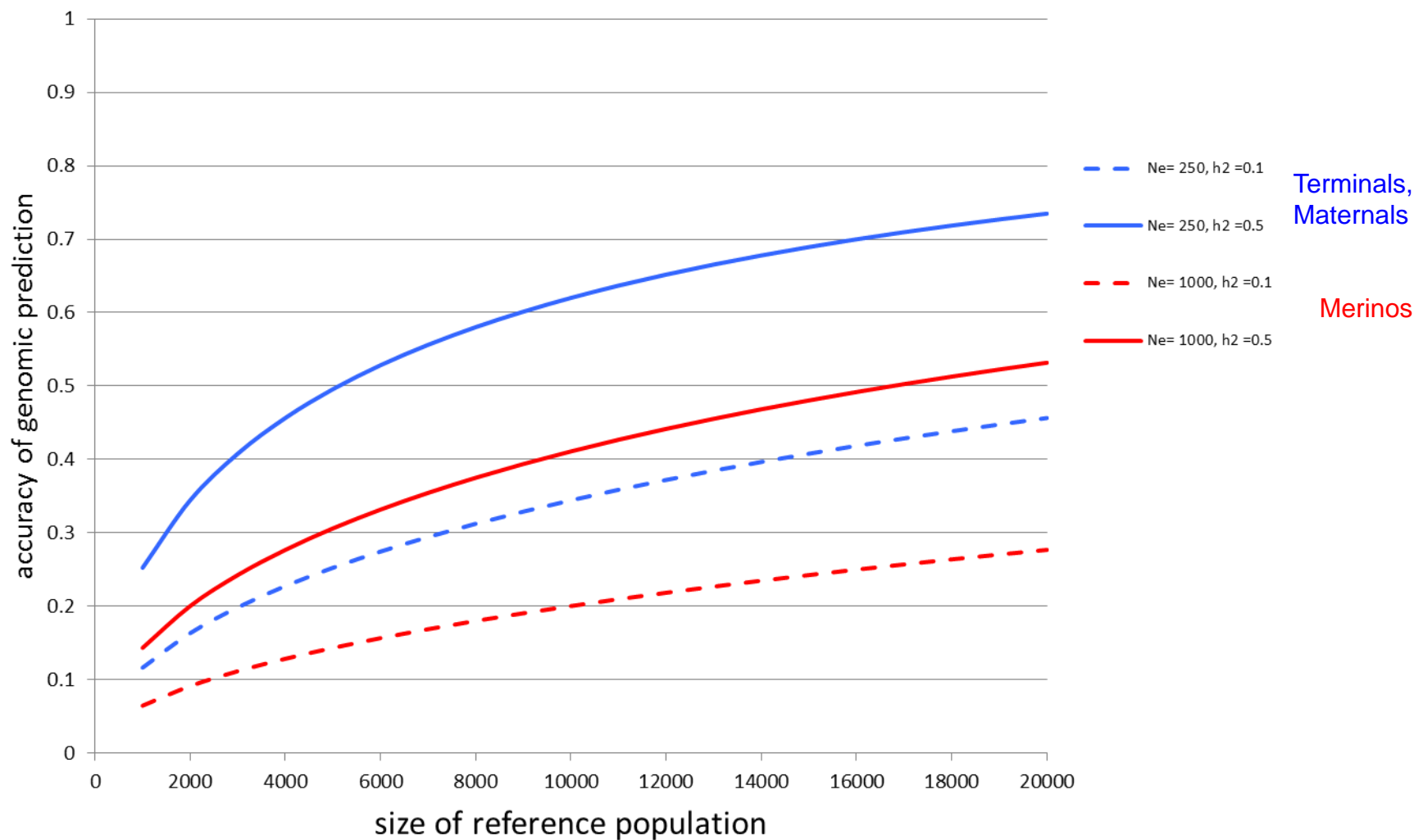# Design of reference populations

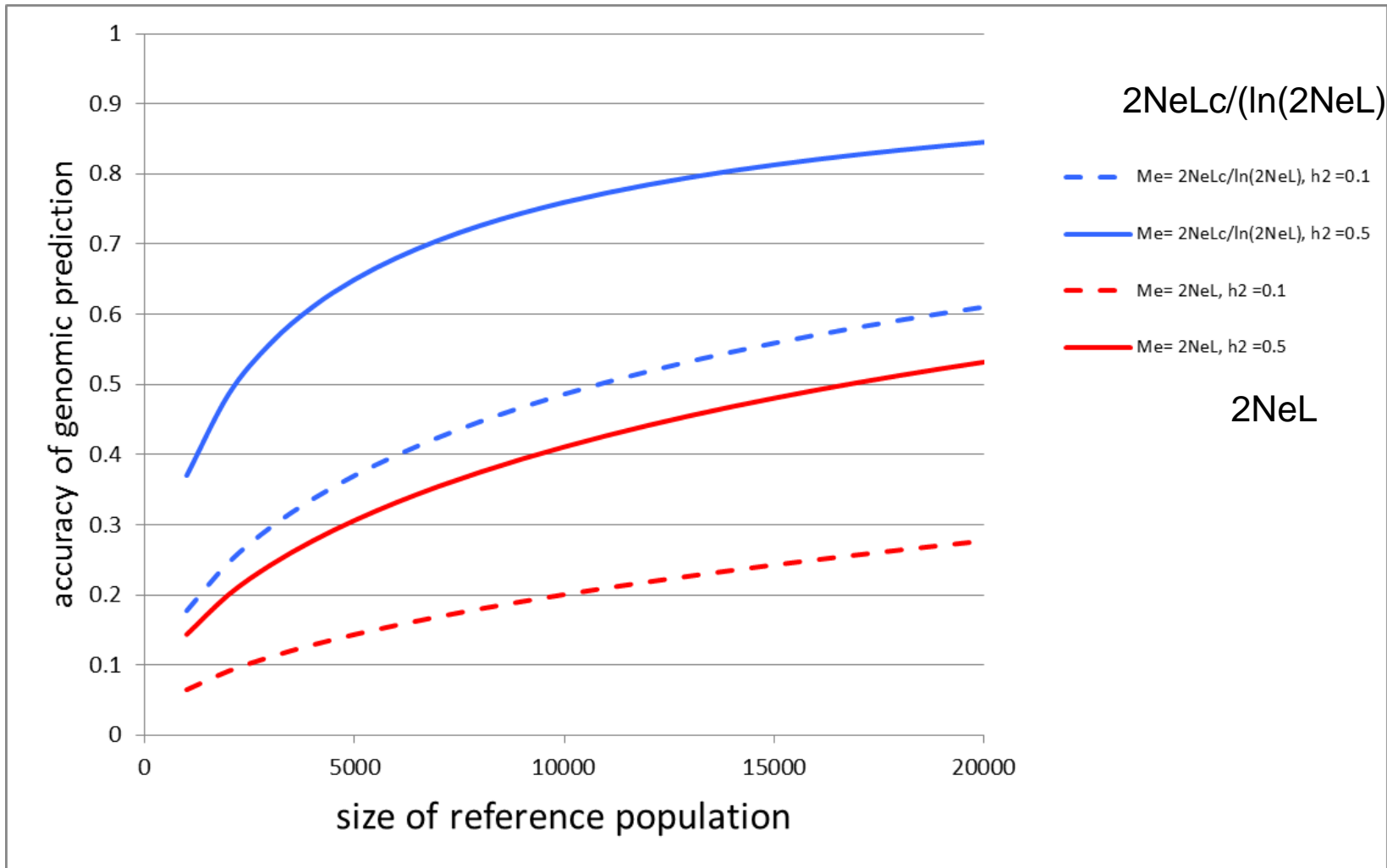size

diversity

relatedness
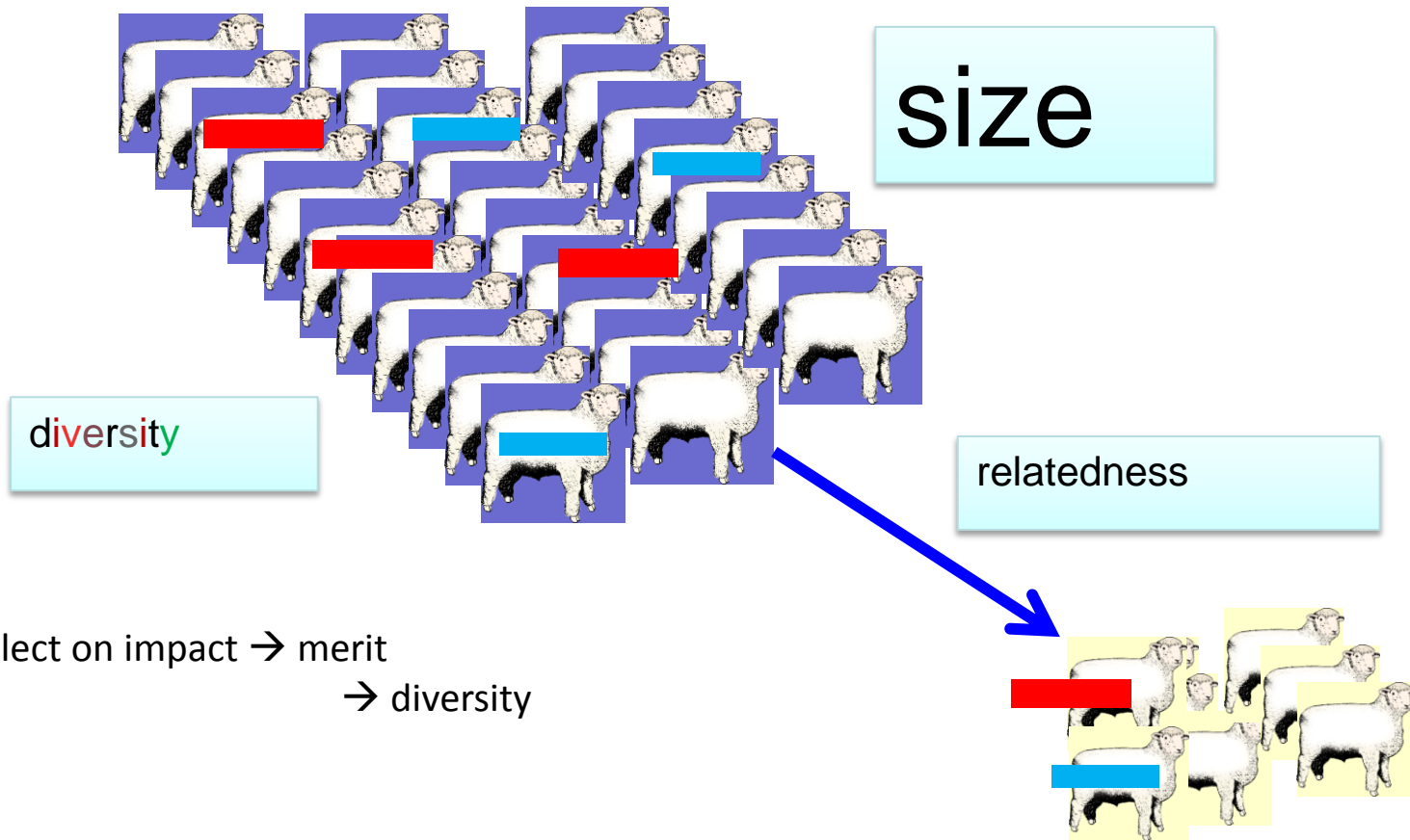
Select on impact → merit
→ diversity

Multi-breed
Across breed?
Longevity of RefPop?

Accuracy of genomic prediction depending on size of reference population
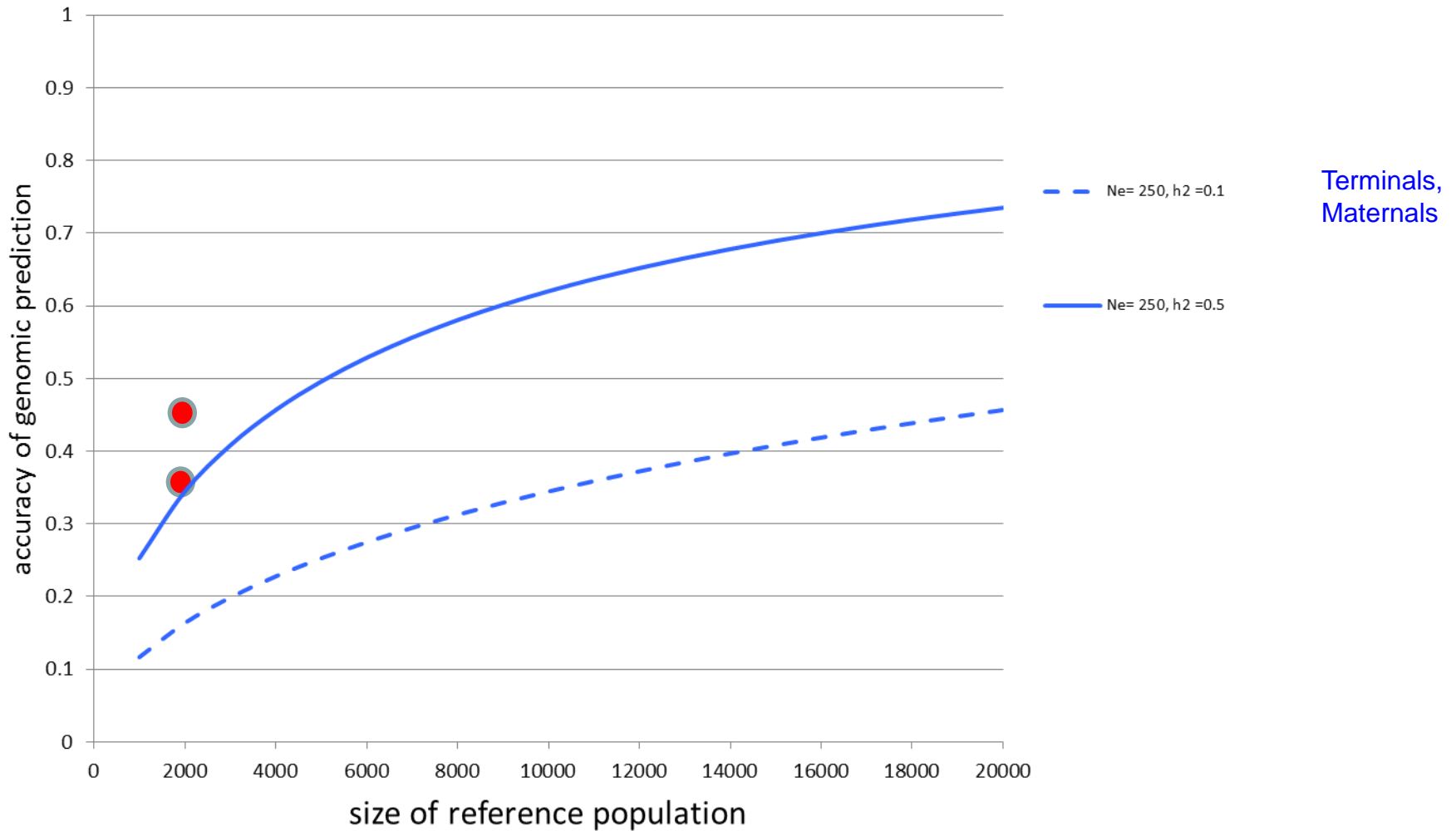
Goddard 2009

une

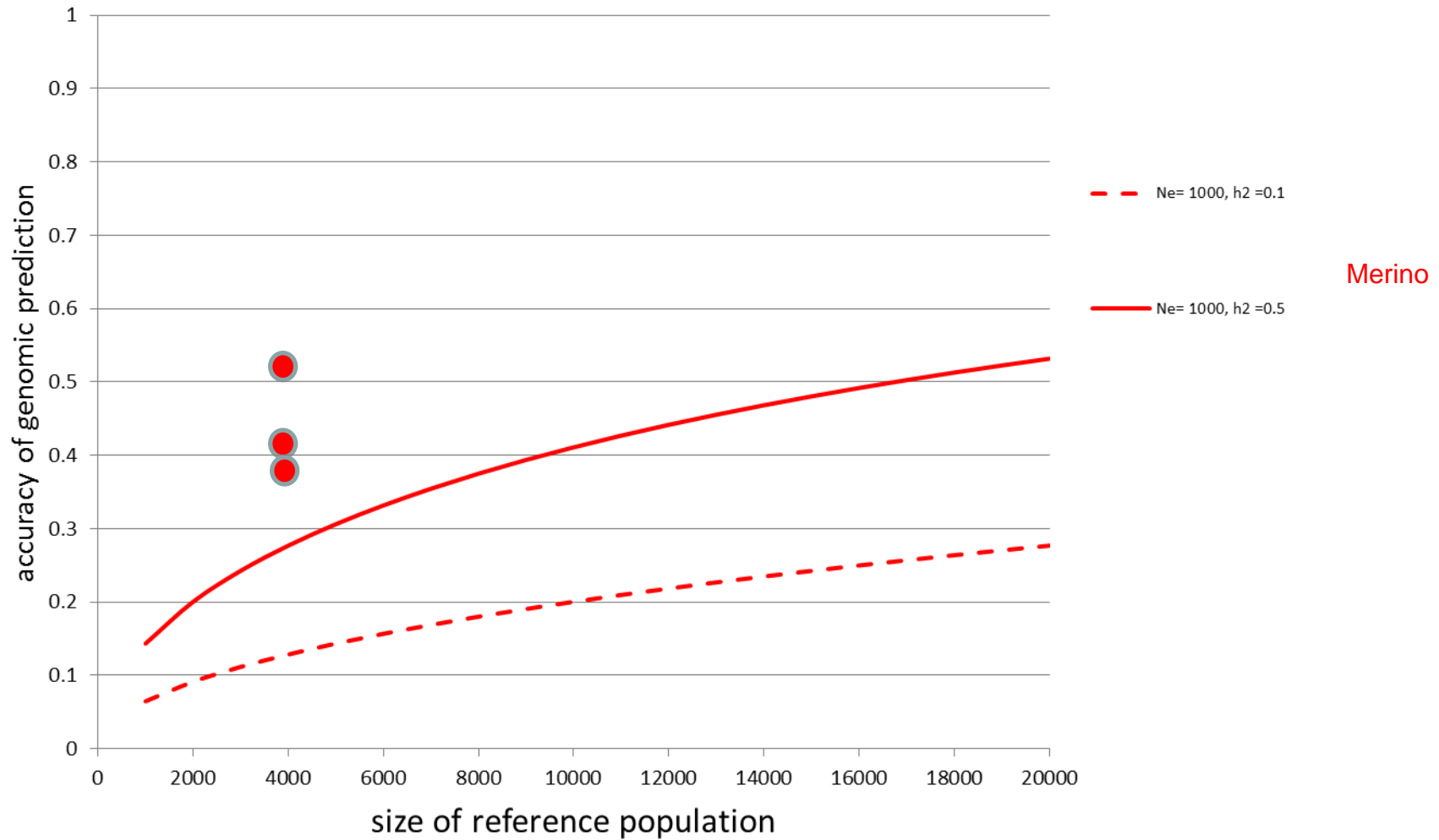# Accuracy, depending on how Me is approximated

design of reference population

- Relatedness between reference population and selection candidates
- Across breeds or lines?
- Number of sires, nr of progeny per sire, which dams?
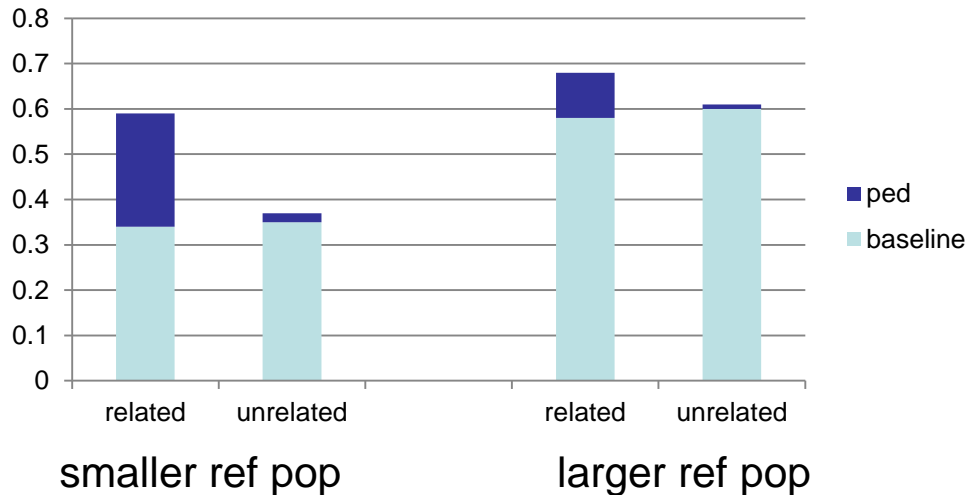
Realized accuracy 1

Realized accuracy 2

Sources of information contributing to GBV accuracy

	BLU P	GBLUP	

half life
1. Variation between families   ++ ++   1 gen

2. Variation within families   0   +   1 gen

3. Markers tracking effects of genome segments/LD 0   ++.+ several gen's
   Info on 'unrelated'

Depending on size of reference population

Reference Pop: How many are needed?

%V$_A$ explained by GBV

Breed	merino	WS, PD	BL
Ne	1000	250	100
Size of reference pop'n	30,000	10,000	5,000
Progeny measured per year[1]	3750	1250	625
h2=0.1	0.33	0.34	0.35
h2=0.3	0.51	0.53	0.54
h2=0.5	0.60	0.62	0.63
Predicted benefit in dG	40%	20%	?

$\cong h^2$

assuming the reference population is 'refreshed' every 8 years

une

Reference Pop: How many are needed?

%V$_A$ explained by GBV

Breed	merino	WS, PD	BL
Size of reference pop'n	12,000	4,000	2,000
Progeny measured per year[1]	1500	500	250
h2=0.1	0.22	0.23	0.23
h2=0.3	0.36	0.37	0.38
h2=0.5	0.44	0.46	0.47
Predicted benefit in dG	20%	10%	?

$\cong \frac{1}{2} h^2$

assuming the reference population is 'refreshed' every 8 years

une