Using genetic regressions for genomic preselection effects

Pete Sullivan (Lactanet, Canada)
Esa Mäntysaari (Luke, Finland)
Gerben deJong (CRV u.a., Netherlands)
Simone Savoia (Interbull, Sweden)

GPS effects for AI bulls

- Single-step GEBV with genotypes partitions properly:
  - **PA** include **between-family** selection effects
  - **MS** include **within-family** selection effects (GPS)
- Pedigree-based EBV without genotypes ignores GPS:
  - true GPS effects in daughter phenotypes are not fully credited as “within-family selection effects” in sire MS
  - ***Sire’s MS is under-predicted***, with offsetting biases (i.e. over-predictions) of sire’s *PA, mates* and/or daughter *environment* effects

GPS-AI bulls in MACE

- ***MACE requires*** pedigree-based EBV without genotypes (most or all are ignoring GPS effects).
  - *Requiring input EBV for MACE that are biased…*
  - But the amount of bias and impacts on MACE are unclear
  - Accounting for GPS effects should reduce GPS-bias

Objective: Develop a **GPS-MACE** model that accounts for GPS effects on the MS-distributions of GPS-AI sires

GPS-MACE model

GPS effects

- Current MACE: $y = \mu + Q_1 g_1 + \mathbf{a} + e$
- Current MACE: $y = \mu + (Q_1 g_1 + \mathbf{PA}) + (\mathbf{MS}) + e$
- **GPS-MACE:** $y = \mu + (Q_1 g_1 + \mathbf{PA}) + (\mathbf{Q_2 s} + \mathbf{ms}) + e$

GPS effects

$Q_1 g_1$ = Base-generation selection

$\mathbf{PA}$ = Between-family selection

$\mathbf{Q_2 s}$ = Within-family pre-selection

$\overline{\mathbf{MS}} = \mathbf{Q_2 s}, \quad \overline{\mathbf{ms}} = \mathbf{0}$

GPS-MACE equations

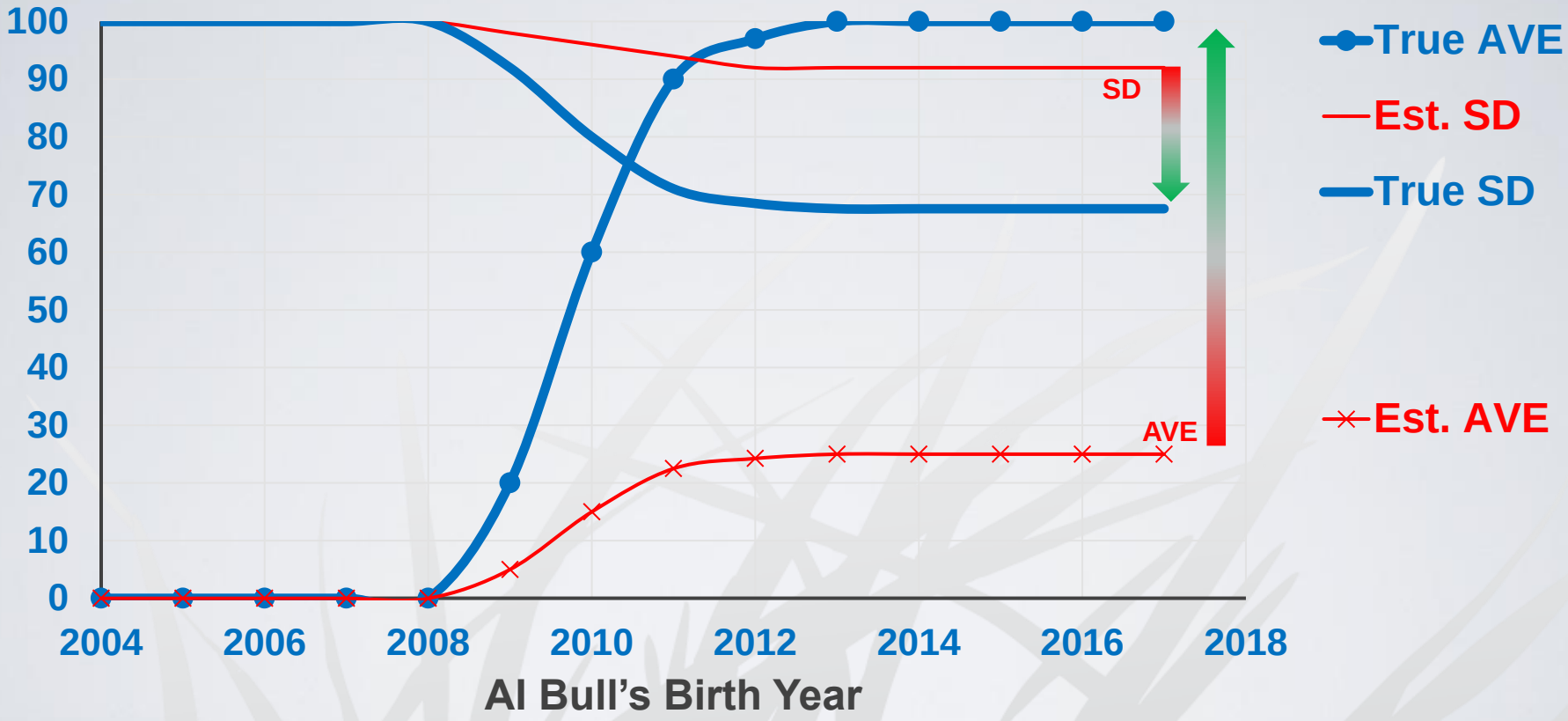
(No V(MS) adjustments yet … but working on it)

$$\begin{bmatrix} \mathbf{X}'\mathbf{D}\mathbf{X} & \mathbf{X}'\mathbf{D}\mathbf{Z} & \mathbf{X}'\mathbf{D}\mathbf{Z}\mathbf{Q}_2 \\ \mathbf{Z}'\mathbf{D}\mathbf{X} & \mathbf{Z}'\mathbf{D}\mathbf{Z} + \mathbf{W} \otimes \mathbf{G}_t^{-1} & \mathbf{Z}'\mathbf{D}\mathbf{Z}\mathbf{Q}_2 \\ \mathbf{Q}_2{}'\mathbf{Z}'\mathbf{D}\mathbf{X} & \mathbf{Q}_2{}'\mathbf{Z}\mathbf{D}\mathbf{Z} & \mathbf{Q}_2{}'\mathbf{Z}'\mathbf{D}\mathbf{Z}\mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \mathbf{Q}_1\mathbf{g} + \mathbf{a} \\ \boldsymbol{s} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}'\mathbf{D}\mathbf{y} \\ \mathbf{Z}'\mathbf{D}\mathbf{y} \\ \mathbf{Q}_2{}'\mathbf{Z}'\mathbf{D}\mathbf{y} \end{bmatrix}$$

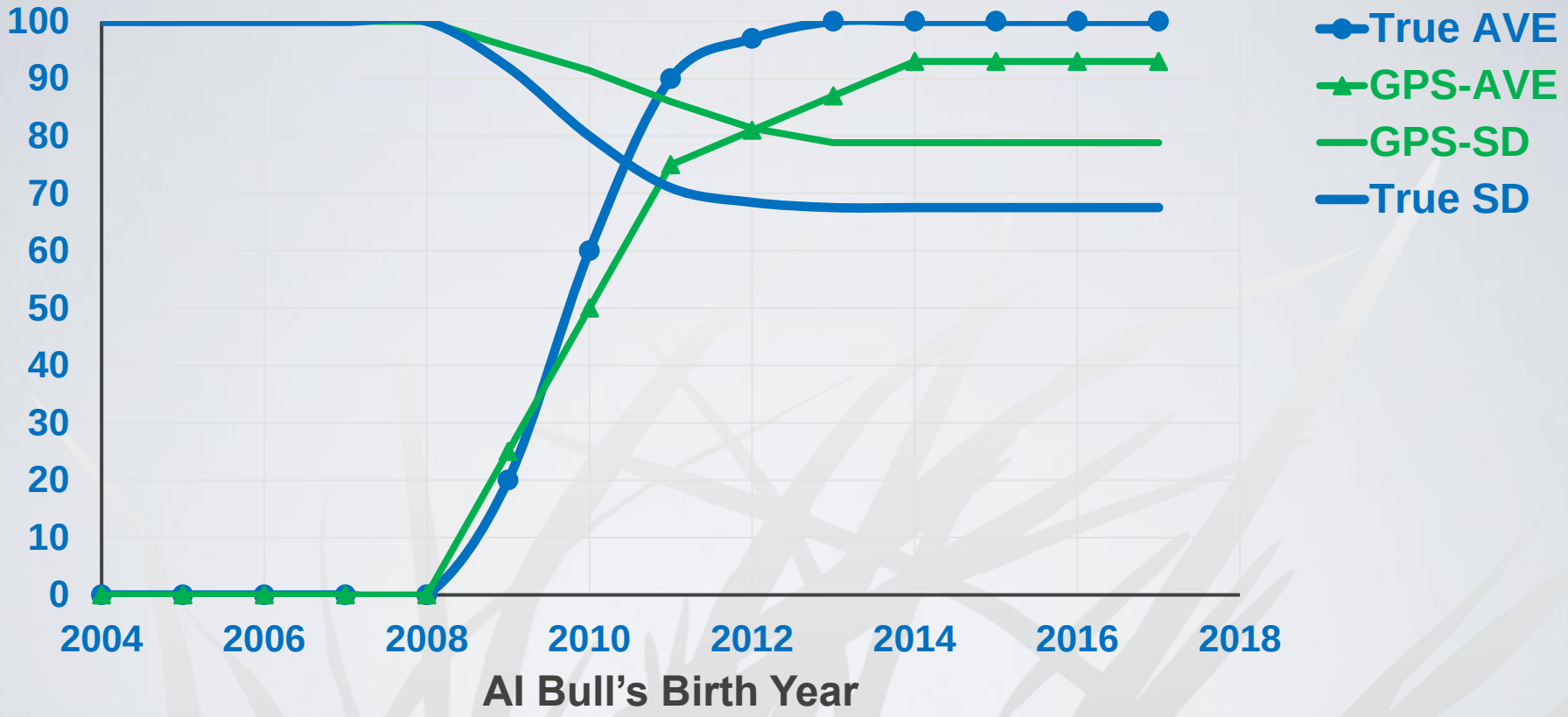
MS Trends (% of maximums)

MACE Estimates of MS



MS Trends (% of maximums)

GPS-MACE Estimates of MS

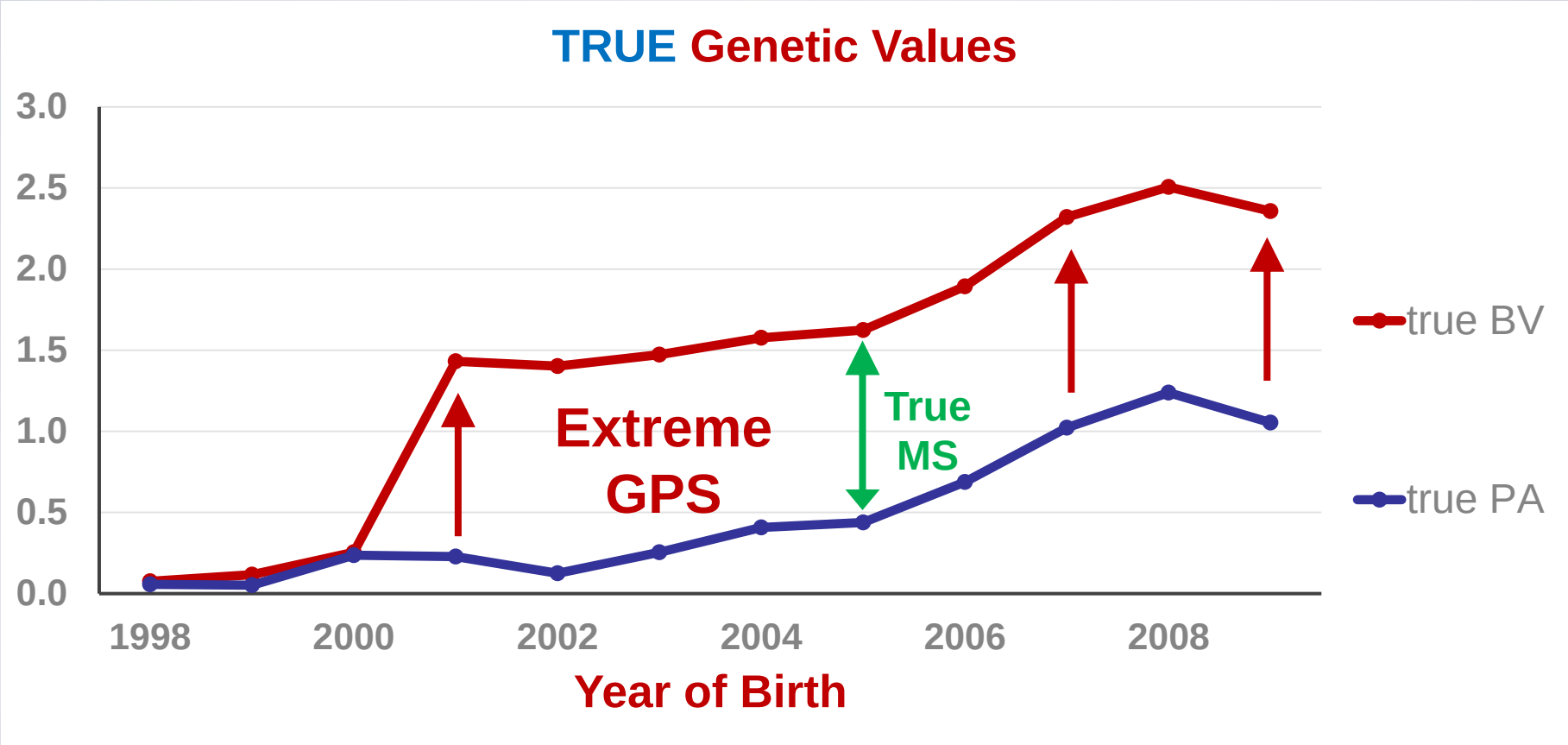


Data

1. **Simulation study**: *unbiased national EBV* input for MACE (GPS practiced in 1 country only)
   - *GPS effects included* in the simulated national EBV
   - BUT *how do we get these national EBV in practice?*
2. **Official data study**: *biased national EBV* input used in MACE, after years of GPS in many countries, but the GPS effects are not properly included in the national EBV computed without genotypes

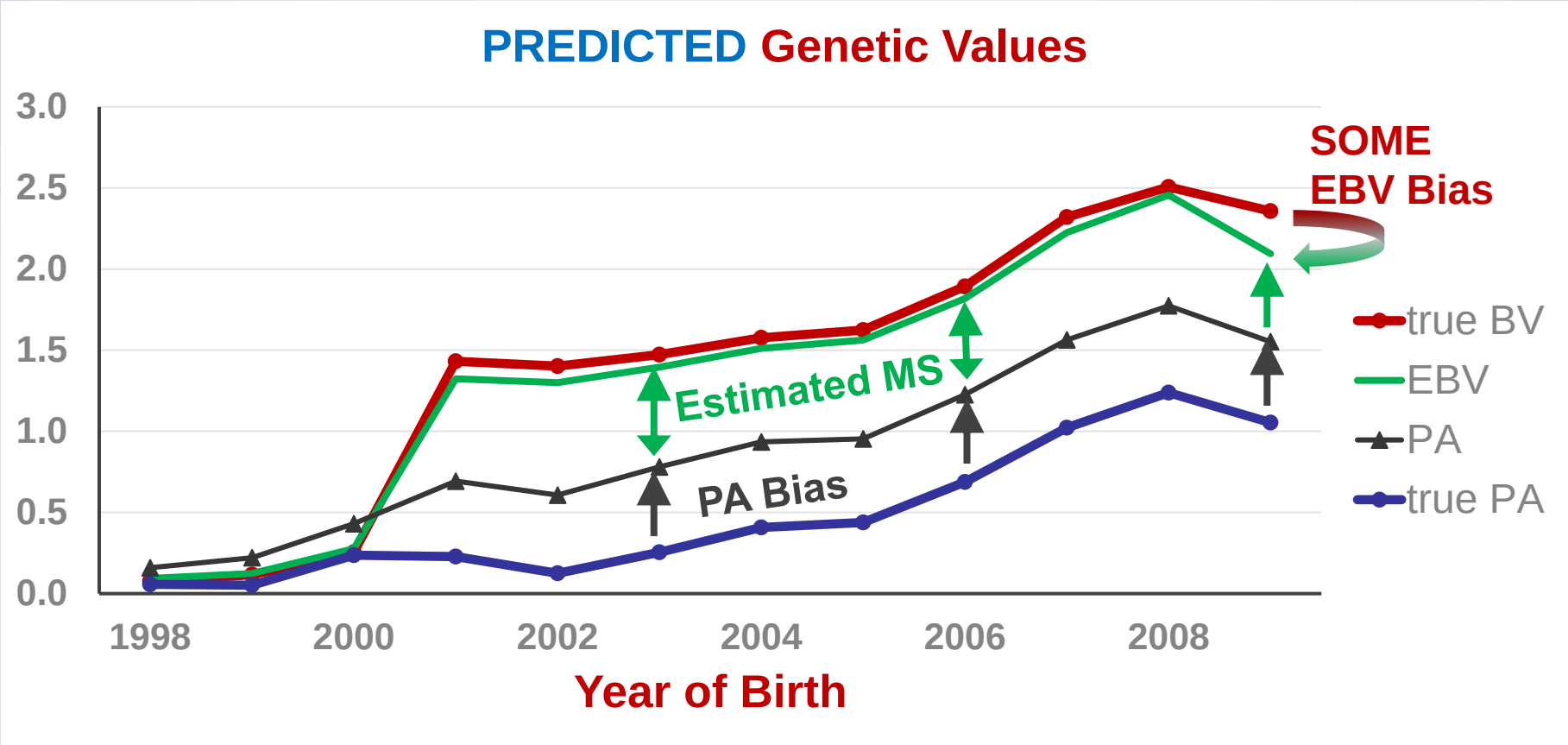
Simulated Data with strong GPS

(Tyrisevä, 2018$_{JDS}$; Benhajali, 2019$_{IB}$)



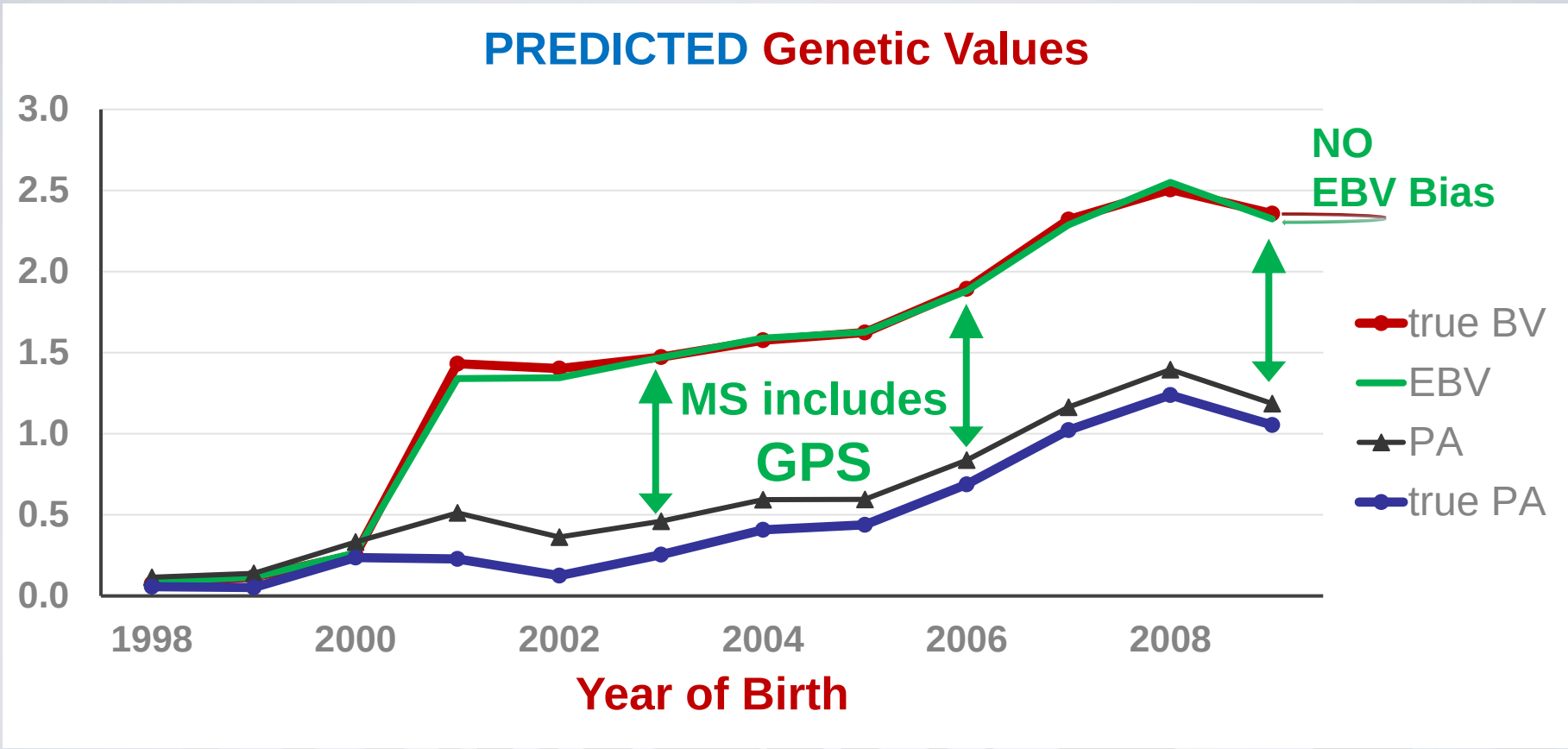
Simulated Data with strong GPS

(MACE with unbiased EBV input)



Simulated Data with strong GPS

(GPS-MACE with unbiased EBV input)



Simulated Data with strong GPS

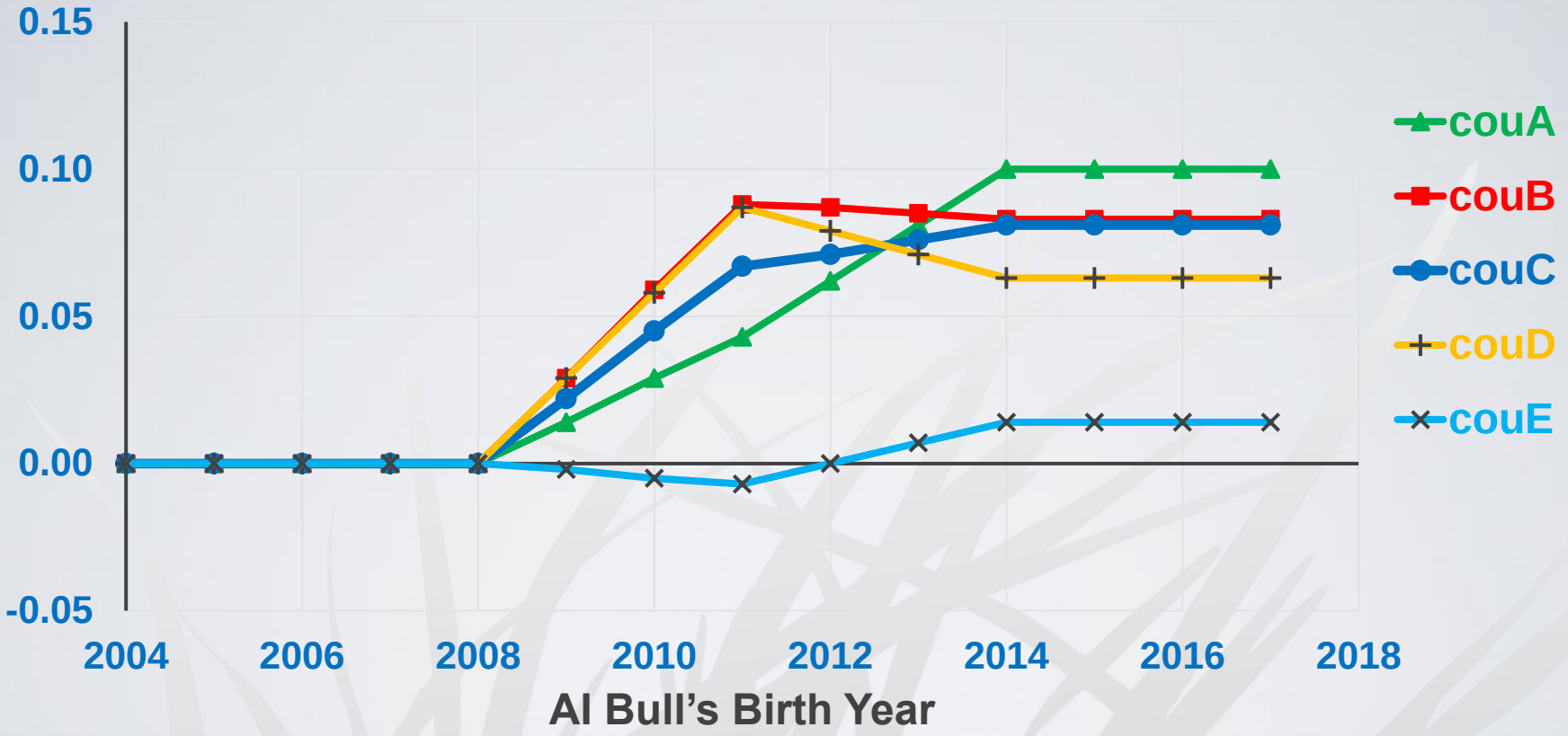
- **IF national EBV are unbiased**, meaning the EBV include GPS effects for recent AI bulls *(not true today)*
  - ➢MACE picks up the GPS effects, but partitions incorrectly
  - ➢MACE proofs of AI bulls (**EBV**) are relatively **GOOD**
    - ***Slightly*** underpredicted ***EBV for only the last AI bull cohort***
  - ➢BUT the **PA** and **MS** predictions are both **WRONG**

EBV = **PA**↑ + **MS**↓

✓***GPS-MACE can fix the problem*** *(IF unbiased national EBV)*

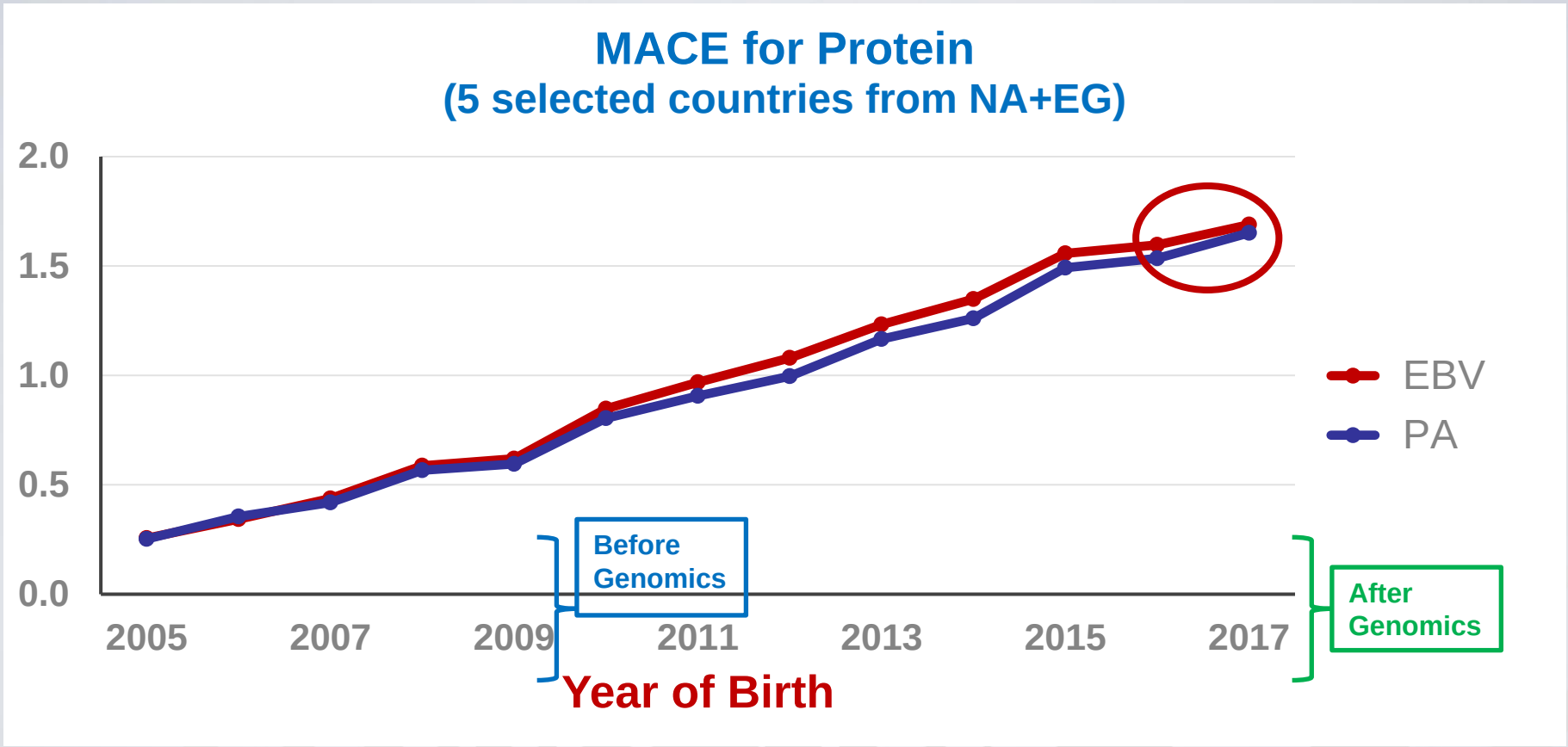
Official data - GPS Trends

($\overline{MS}$ for Protein, April 2022)



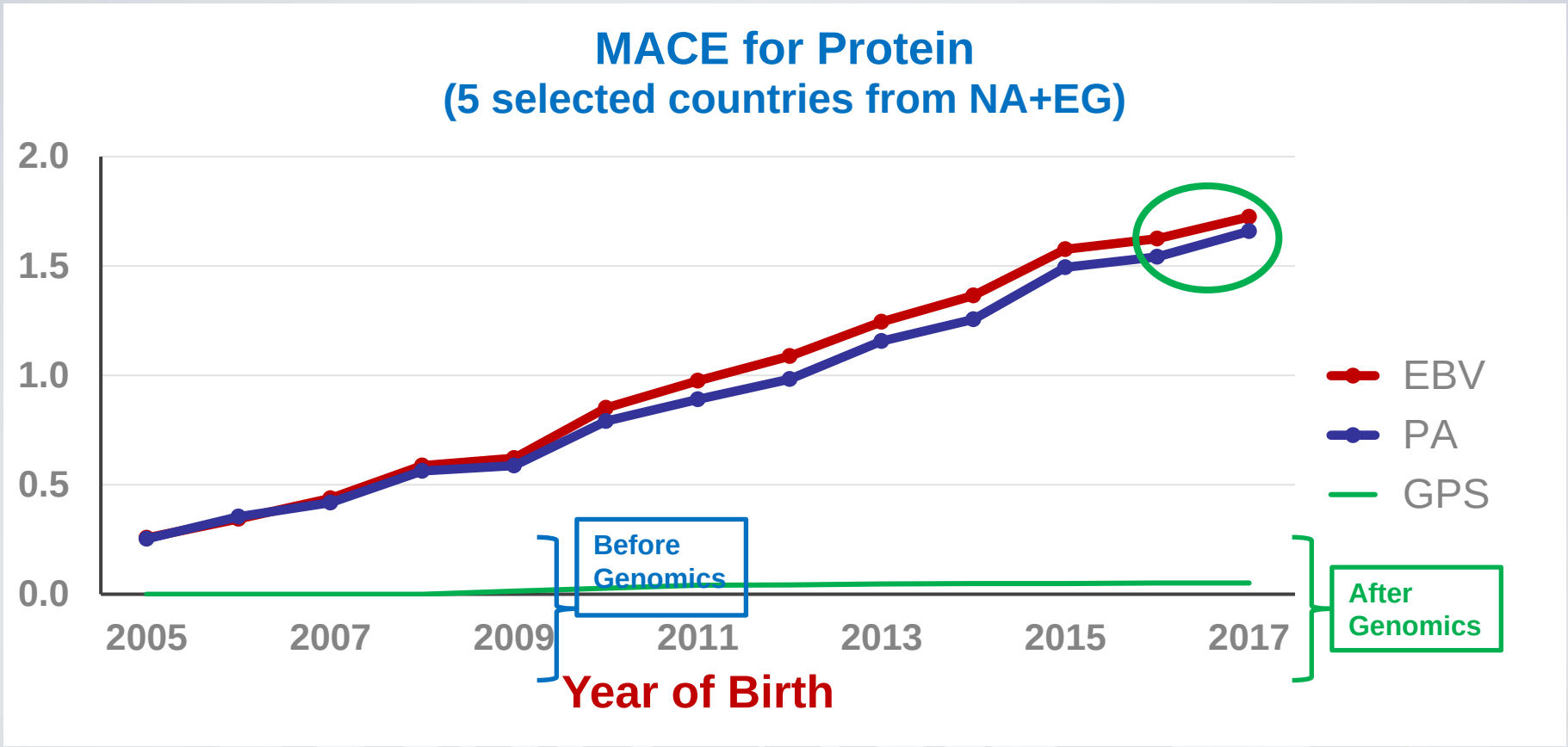
Official data – April 2022

(MACE with biased EBV input)



Official data – April 2022

(GPS-MACE with biased EBV input)



Official Data without genotypes

- MS averages (EBV-PA) from **MACE** were small, even though GPS was practiced in many countries
- MS averages (EBV-PA) from **GPS-MACE** were larger, but still relatively small
- **GPS estimates were generally positive**, for countries with national genomic evaluation programs

Official Data without genotypes

- GPS-MACE improves the PA-MS partition, but it does not "make up" GPS effects being excluded from both the PA and the MS
  - i.e. the GPS effects going incorrectly to the EBV of sires' mates, and environment effects of the daughters
- Much bigger improvements can be expected with GPS-MACE after removing GPS biases at the national level, so that GPS effects are more fully expressed in the MS and hence national EBV of GPS sires feeding into MACE

What's Next

- Continue developing national methods that can generate better MACE input data, which are
  - including GPS effects while excluding genotype effects
  - A GPS-MACE system will become more necessary as improved national EBV become available
- Continue refining GPS-MACE model
  - Still need to reduce V(MS) as f($\overline{\mathbf{MS}}$)
  - Verify reasonable results for all traits and countries
- GPS-MACE versus MACE using new validation tests
  - Focus on future prediction of PA and MS