

Impact of data sub-setting strategies on variance component estimation for Interbeef evaluations

Renzo Bonifazi, J. Vandenplas, J. ten Napel,
A. Cromie, R. F. Veerkamp, M. P. L. Calus





Dr. Roel Veerkamp

Dr. Mario Calus

Dr. Jeremie Vanderplas

Dr. Jan ten Napel



Dr. Ross Evans

Dr. Thierry Pabiou

Dr. Andrew Cromie



Dr. Hossein Jorjani

Dr. Alexis Michenet



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Dr. Martino Cassandro



UNIVERSITÀ
DEGLI STUDI
DI PERUGIA

Dr. Emiliano Lasagna



Dr. Kaarina Matilainen

International genetic evaluations for beef cattle



- **Role of Interbeef:** comparison of animals' genetic values
- Exchange of data: phenotypes and (international) pedigree
- **International** Estimated Breeding Values
- 12 countries (10 Populations)
- 5 breeds: Limousin, Charolais, Hereford, Angus, Beef Simmental

International genetic evaluations for beef cattle



- Evaluated **traits**:
 - Age adjusted weaning weight
 - Calving traits (calving ease and birth weight)
- **Interbeef is growing**: new countries and traits (carcass quality, fertility)
- Estimation of **across-countries genetic correlations**:
 - New trait and countries
 - Changes in the model definition (national and international)

Estimation of across-countries genetic correlations (r_g)

- Genetic connections → **Common Bulls** (CB)
- Low usage of Artificial Insemination → **low genetic connections**
- Traits with **maternal** genetic effect

Current procedure:

- Sire-MGS bi-variate model (data amount)
- Estimation challenges (convergence, default r_g)

Aim of the study

Estimation of across-countries r_g using a **multi-trait approach** and investigation of the **impact of data sub-setting** strategies

Model

Multi-Trait Animal Model (countries as traits):

- National **fixed** and **random** environmental and genetic effects

$$y = Xb + Cr + Zu + Wm + Ppe + e$$

b and **r** = Fixed and Random Environmental

u and **m** = Random direct (animal) and maternal genetic

pe = Random permanent environment (dam)

e = Residual

- Covariances between countries = 0, beside direct and maternal genetic effects

Data – Age adjusted weaning-weight (AWW)

Limousin

8 populations

January 2018 evaluation

POP	Total	%	Herds	Year of Birth
CZE	10,500	0.34	121	1991 – 2017
DFS	90,456	2.90	9,190	1980 – 2017
ESP	33,152	1.06	188	1989 – 2011
GBR	127,840	4.10	745	1972 – 2017
IRL	20,609	0.66	1,304	1975 – 2017
FRA	2,714,368	87.12	6,677	1972 – 2017
DEU	88,628	2.84	881	1981 – 2017
CHE	30,045	0.96	224	1993 – 2017
Total	3,115,598	100	19,330	1972 - 2017



Assessment of genetic connections

	CZE	DFS	ESP	GBR	IRL	FRA	DEU	CHE
CZE	554							
DFS	65	4,375						
ESP	44	76	1,188					
GBR	67	109	97	5,486				
IRL	64	94	78	239	2,073			
FRA	157	171	358	396	200	57,784		
DEU	101	143	105	125	120	339	4,366	
CHE	63	73	71	72	65	342	188	1,699

Diagonal: Number of national sires

Lower diagonal: Number of Common Bulls

Assessment of genetic connections

	CZE	DFS	ESP	GBR	IRL	FRA	DEU	CHE
CZE	554	0.06	0.07	0.04	0.14	0.11	0.06	0.12
DFS	65	4,375	0.06	0.05	0.07	0.13	0.06	0.06
ESP	44	76	1,188	0.04	0.12	0.13	0.06	0.08
GBR	67	109	97	5,486	0.15	0.13	0.04	0.04
IRL	64	94	78	239	2,073	0.12	0.07	0.10
FRA	157	171	358	396	200	57,784	0.15	0.13
DEU	101	143	105	125	120	339	4,366	0.11
CHE	63	73	71	72	65	342	188	1,699

$$GS_{POP_1, POP_2} = \frac{N. Off. \text{ Common Bulls}}{N. Off. \text{ Total bulls}}$$

Rekaya R. et al., 1999

Low

GS < 0.05

Medium

0.05 < GS < 0.10

High

GS > 0.10

Scenario tested

- **All data** (3.1 million data)

Data subsets: ~ 0.5 million

French population reduction by selection of **FRA herds:**

- **Random:** 3 samples (average)
- **GS (CB):** Genetic Similarity at herd level (Common Bulls)
- **GS (CB + CMGS):** Genetic Similarity at herd level (CB + Common MGS)

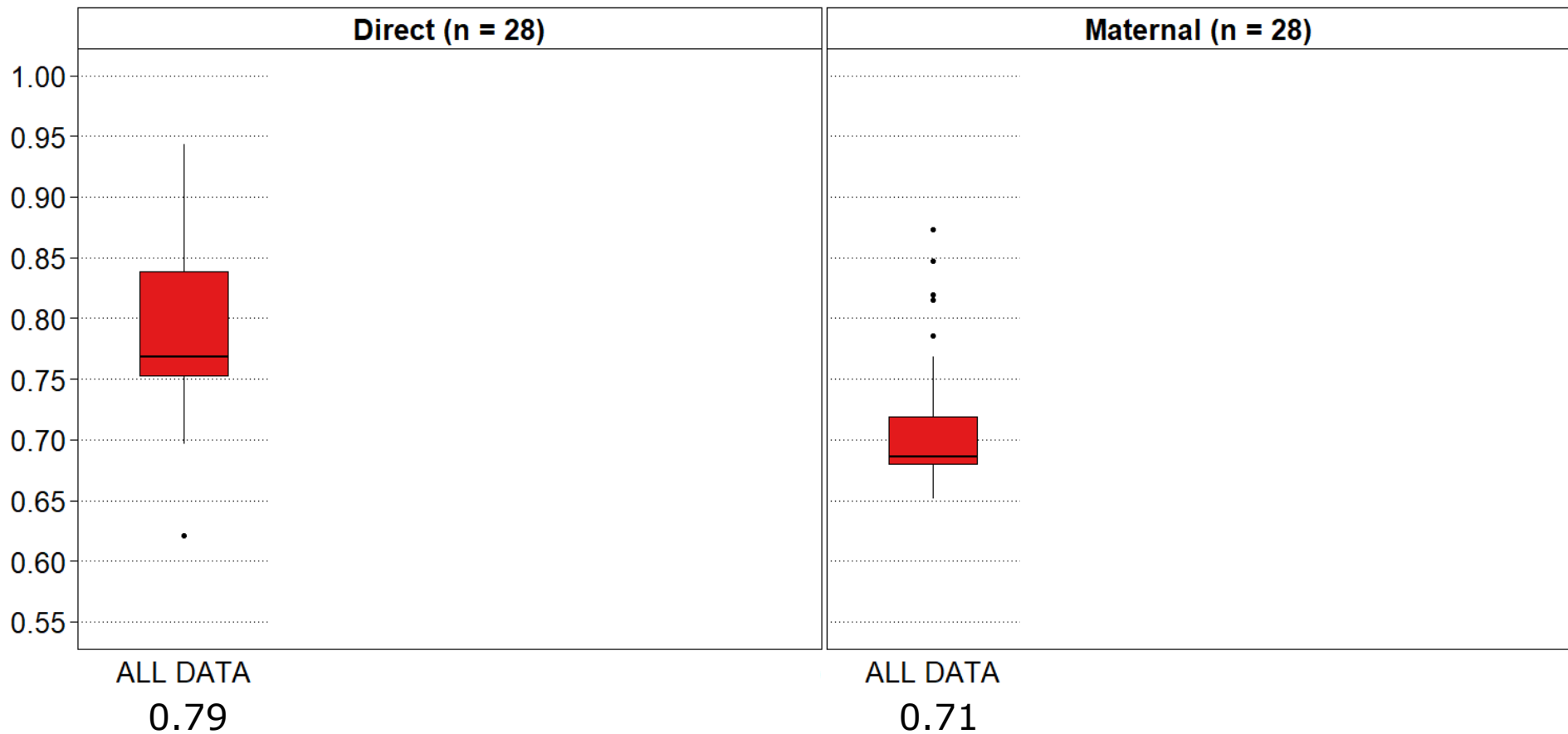
MiX99 – VCE (MC EM REML¹)

Most recent estimates as starting values

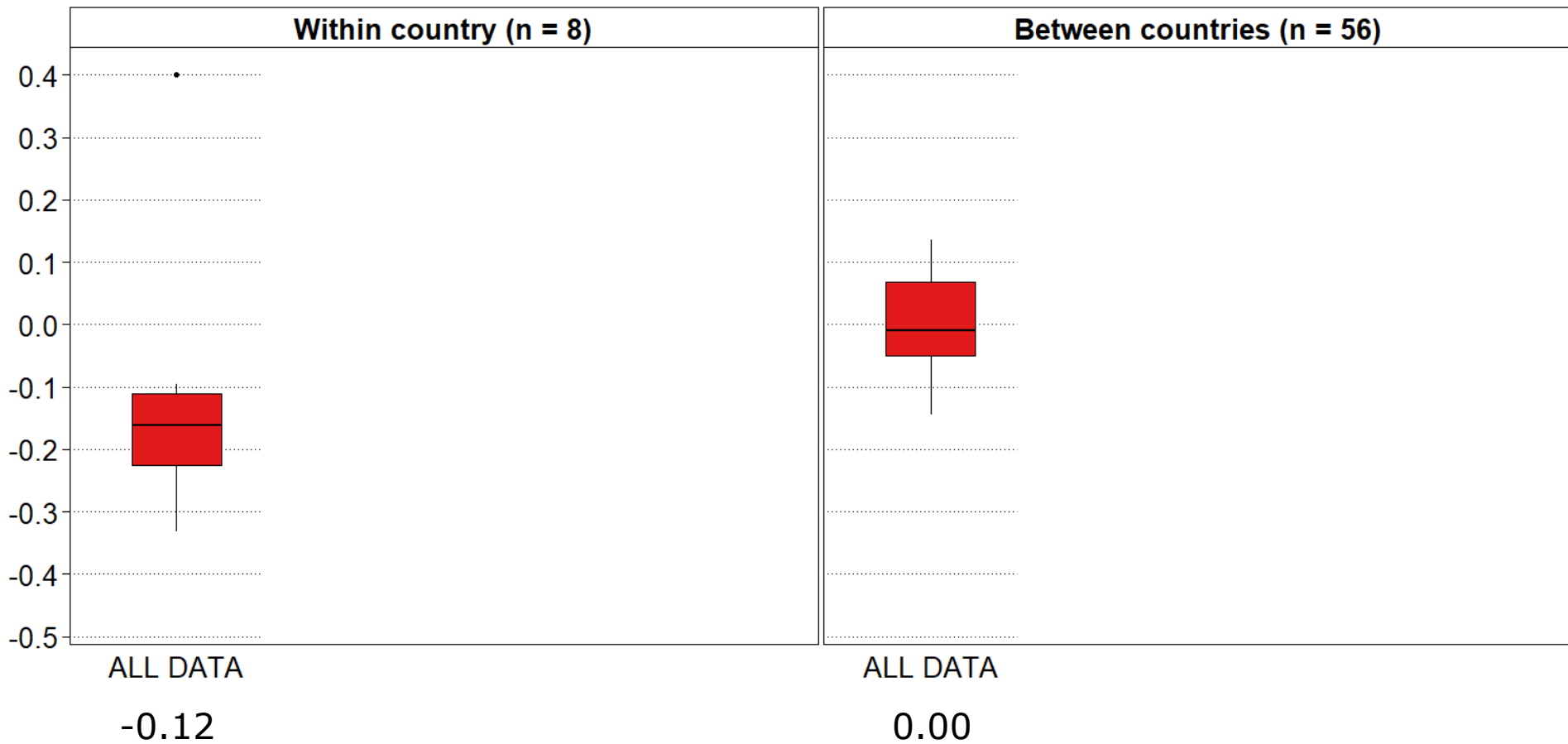
All data r_g (lower diag.) and Standard Errors of r_g (upper diag.)

		DIRECT							MATERNAL								
		CZE	DFS	ESP	GBR	IRL	FRA	DEU	CHE	CZE	DFS	ESP	GBR	IRL	FRA	DEU	CHE
DIRECT	CZE		0.16	0.21	0.15	0.19	0.12	0.14	0.22	0.10	0.15	0.20	0.16	0.16	0.10	0.12	0.19
	DFS	0.87		0.16	0.10	0.13	0.07	0.10	0.15	0.13	0.06	0.14	0.11	0.12	0.06	0.09	0.16
	ESP	0.74	0.77		0.17	0.20	0.14	0.17	0.22	0.20	0.18	0.16	0.16	0.16	0.10	0.15	0.22
	GBR	0.71	0.82	0.94		0.14	0.06	0.10	0.18	0.15	0.12	0.15	0.06	0.11	0.06	0.11	0.18
	IRL	0.83	0.76	0.87	0.91		0.11	0.13	0.21	0.16	0.15	0.18	0.11	0.14	0.09	0.12	0.20
	FRA	0.76	0.89	0.77	0.82	0.76		0.06	0.13	0.09	0.08	0.10	0.08	0.10	0.02	0.06	0.12
	DEU	0.76	0.94	0.76	0.77	0.62	0.81		0.14	0.13	0.11	0.16	0.13	0.13	0.06	0.05	0.15
	CHE	0.85	0.81	0.76	0.71	0.70	0.70	0.70		0.19	0.18	0.23	0.20	0.19	0.11	0.14	0.10
MATERNAL	CZE	-0.12	0.04	0.07	0.12	-0.01	-0.10	0.08	0.01		0.20	0.26	0.21	0.22	0.14	0.16	0.27
	DFS	-0.05	-0.14	0.02	-0.01	-0.02	-0.11	-0.07	-0.01	0.68		0.23	0.18	0.19	0.11	0.14	0.24
	ESP	0.03	0.09	-0.22	-0.08	-0.09	-0.05	0.05	0.02	0.67	0.68		0.24	0.25	0.15	0.21	0.33
	GBR	0.14	0.06	-0.03	-0.10	-0.03	-0.14	0.07	0.08	0.79	0.69	0.70		0.18	0.12	0.16	0.26
	IRL	-0.03	0.07	-0.06	-0.05	-0.19	-0.12	0.12	0.11	0.69	0.68	0.81	0.72		0.16	0.16	0.24
	FRA	-0.02	-0.05	-0.03	-0.06	-0.09	-0.33	-0.01	0.08	0.85	0.69	0.71	0.87	0.82		0.07	0.17
	DEU	-0.02	-0.09	-0.03	-0.01	0.06	-0.10	-0.24	0.09	0.68	0.68	0.67	0.69	0.68	0.69		0.20
	CHE	0.12	0.11	0.07	0.08	0.03	-0.05	0.06	0.40	0.73	0.68	0.67	0.66	0.65	0.77	0.66	

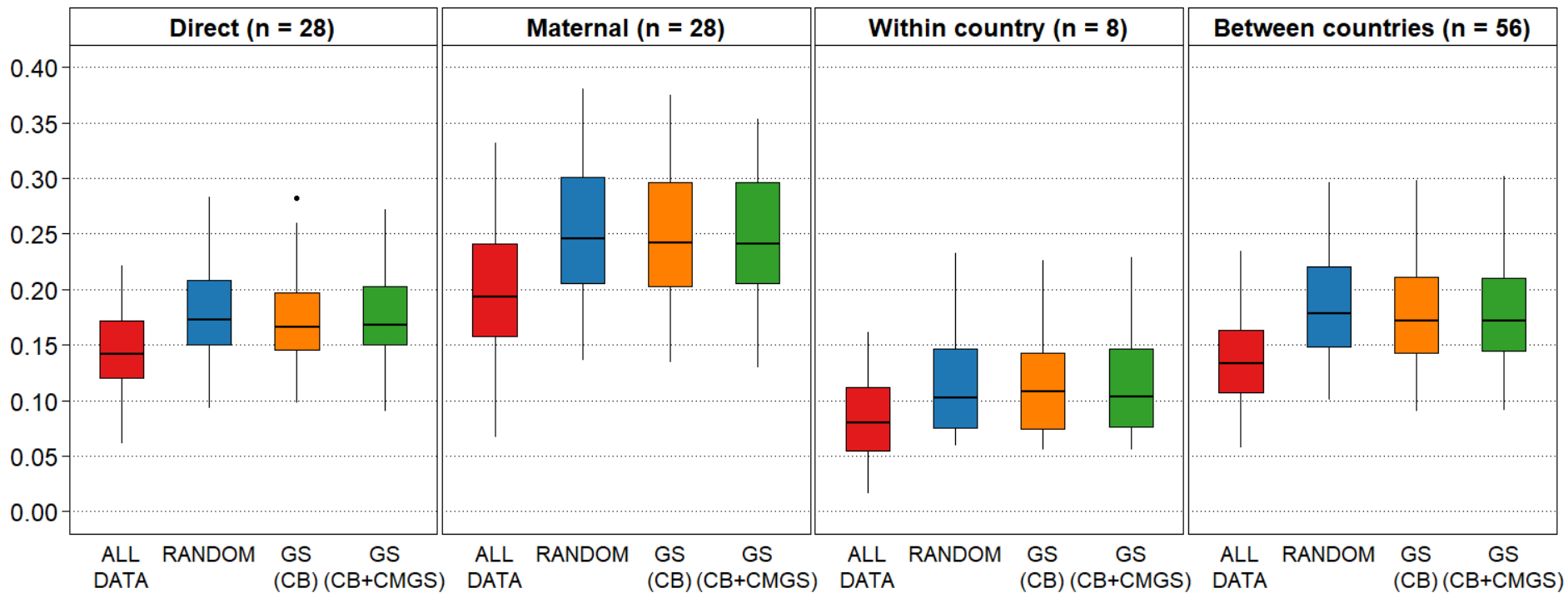
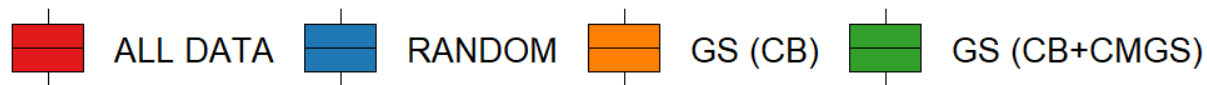
Direct and maternal r_g



Direct-maternal r_g



Standard Errors of r_g



Conclusions

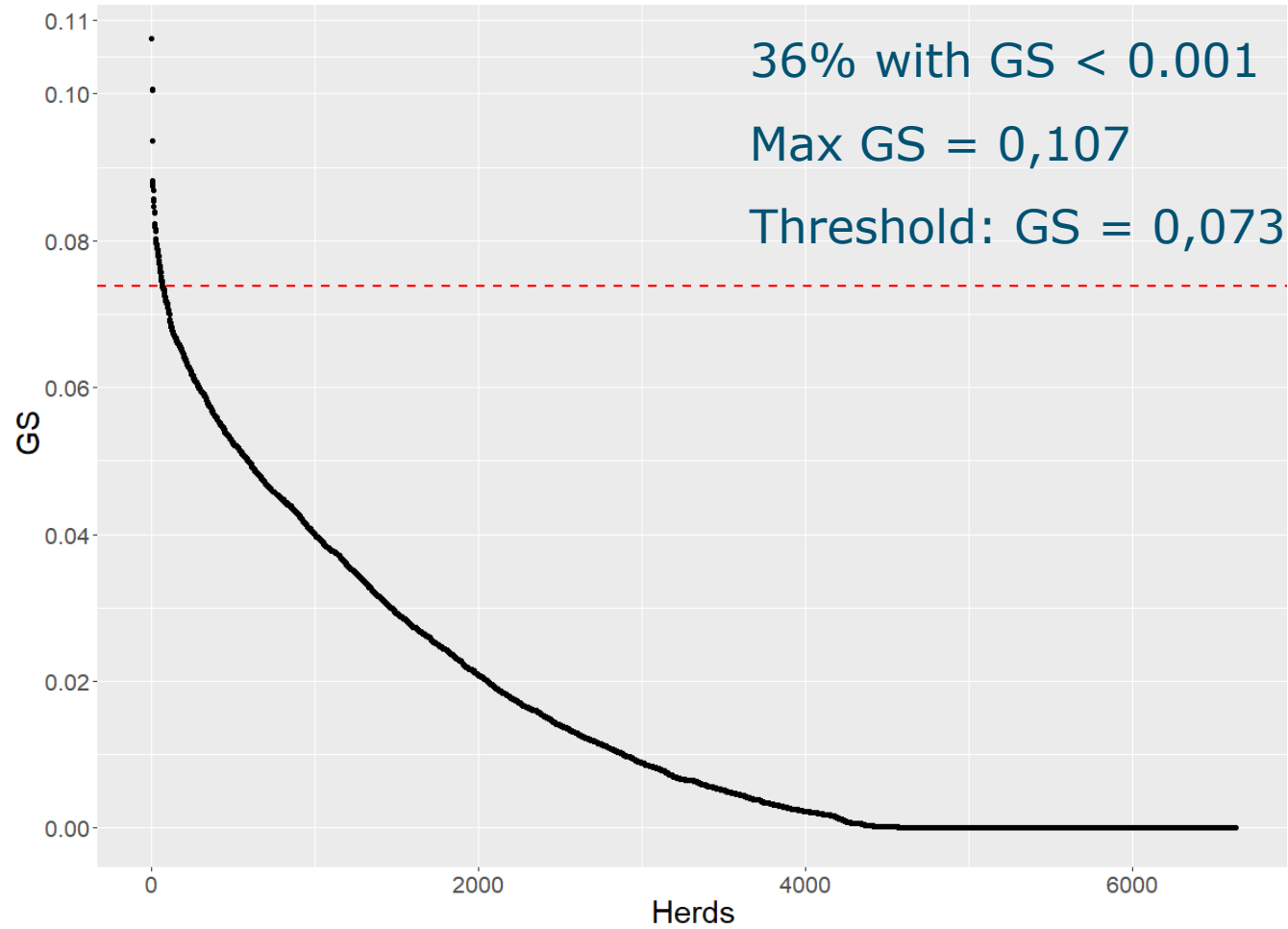
- Estimation of r_g with multi-trait approach is feasible, but requires long computational time: 1 month (3 million records)
- Data sub-setting approaches are required: 2 weeks (0.5 m. records)
- No sensible differences across scenarios:
 - Smaller r_g : -0.02 direct and -0.05 maternal
 - Biggest differences direct-maternal r_g up to:
 - 0.13 within country and - 0.12 between countries
 - Higher Standard Errors (+ 0.02 up to + 0.06)

Thanks for your attention

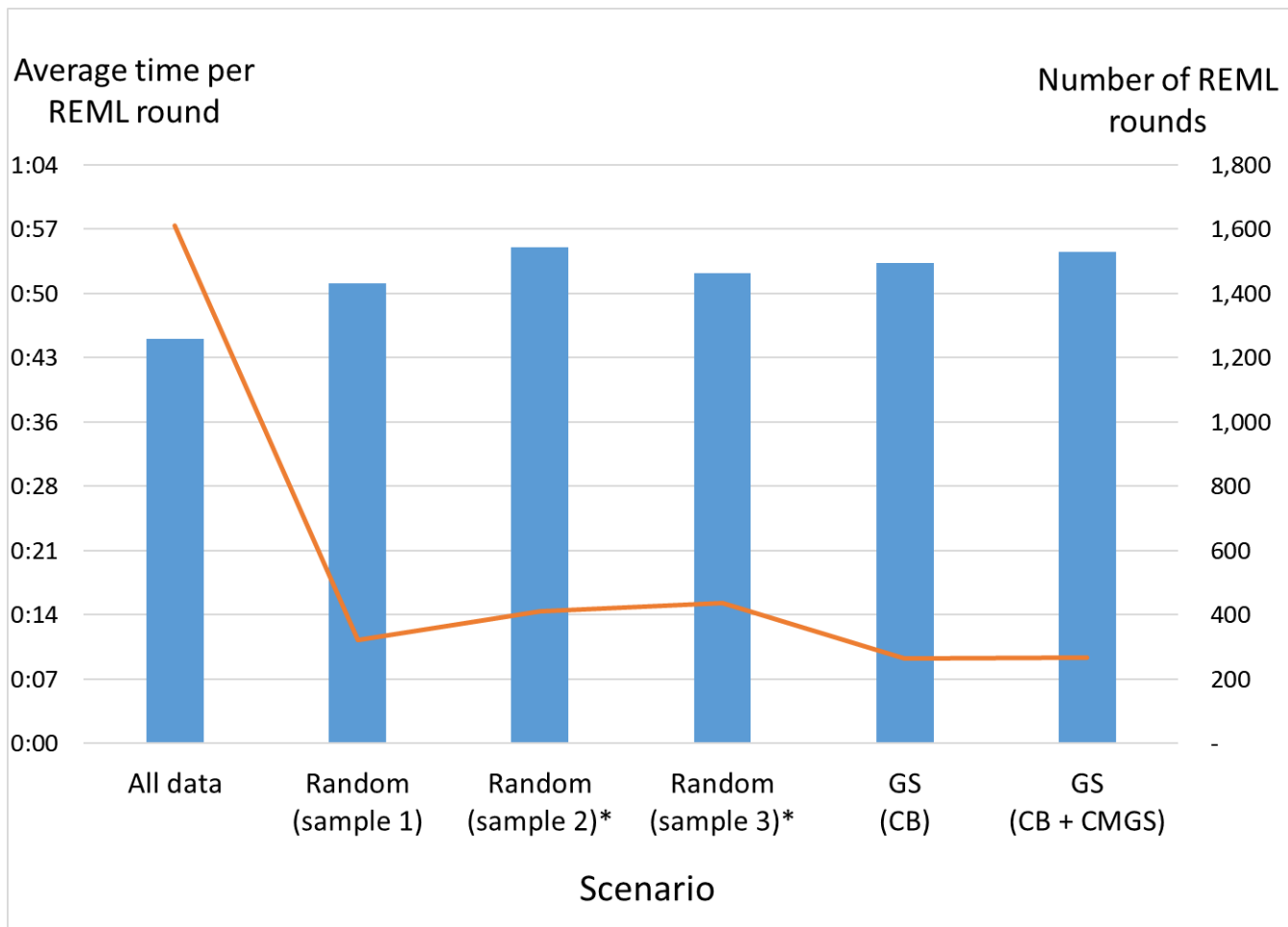


renzo.bonifazi@wur.nl

GS (CB) – GS at the herd level



Computational time

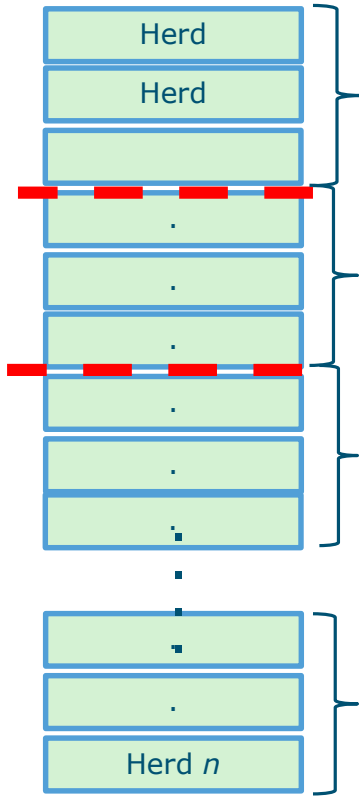


Line = average time per REML round

Bars = Number of REML rounds

CPU Freq: 4.0 GHz
(* 3.6 GHz)

Random. Random selection of FRA herds



Random selection of 5% of the FRA-herds



Retain all data from those herds



20 subsets of data from FRA

Total data (~ 0.5 m)



Analysed 3 subsets out of the 20
(Highest GS for FRA)

Common Maternal Grand Sires (**CMGS**)

- Common MGS can provide genetic links for better estimation of *maternal* and *direct-maternal* genetic covariances.

Common Bulls are 1,441

Common MGS are 3,828.

	CMGS also CB		
Connected countries	Yes	No	Sum
2	554	2,486	3,040
3	204	303	507
4	77	42	119
5	57	15	72
6	42	4	46
7	24	1	25
8	19	0	19
Sum	977	2,851	3,828 ²²

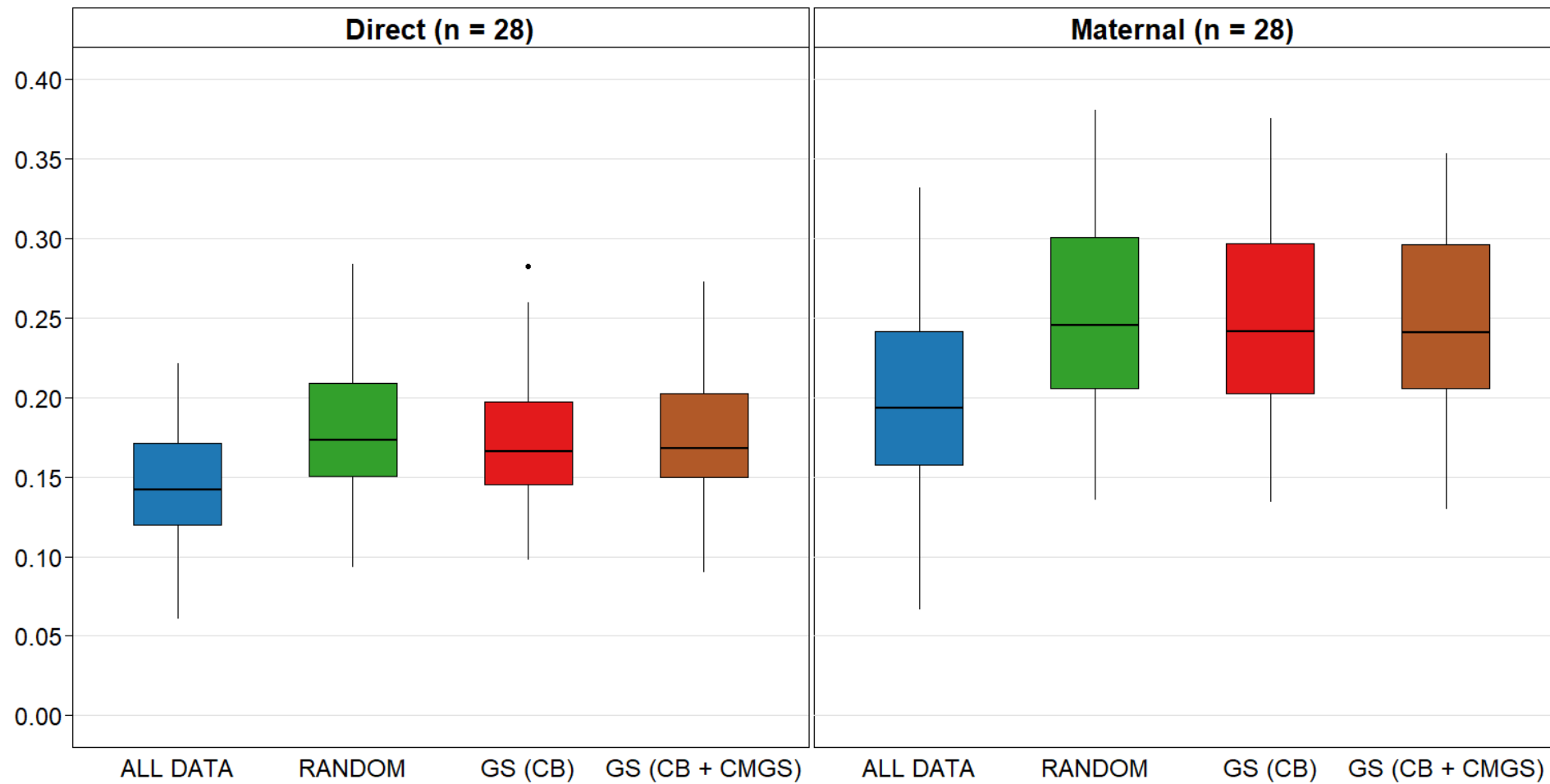
Differences between MiX99 and DMU estimates – 8 POP

		DIRECT								MATERNAL											
		CZE	DFS	ESP	GBR	IRL	FRA	DEU	CHE	CZE	DFS	ESP	GBR	IRL	FRA	DEU	CHE				
DIRECT	CZE																	<i>dir-dir</i>	<i>dir-mat (off diag)</i>		
	DFS	-0.01																average	-0.02	average	0.04
	ESP	-0.02	-0.02															min	-0.03	min	-0.18
	GBR	-0.03	-0.02	-0.01														max	0.00	max	0.11
	IRL	-0.02	-0.02	-0.01	-0.01													<i>dir-mat (diag)</i>	<i>mat-mat</i>		
	FRA	-0.03	-0.02	-0.02	-0.02	-0.03												average	0.04	average	0.00
	DEU	-0.01	0.00	-0.02	-0.02	-0.03	-0.02											min	0.02	min	-0.01
	CHE	-0.01	-0.01	-0.01	-0.02	-0.02	-0.02	-0.01										max	0.06	max	0.02
MATERNAL	CZE	0.06	0.08	0.05	0.06	0.03	0.04	0.08	0.07												
	DFS	0.01	0.02	0.01	0.01	0.00	0.00	-0.18	0.02	0.00											
	ESP	0.04	0.06	0.02	0.03	0.01	0.02	0.07	0.04	0.02	-0.01										
	GBR	0.06	0.09	0.05	0.05	0.04	0.04	0.09	0.06	0.01	0.00	0.01									
	IRL	0.05	0.07	0.04	0.04	0.02	0.03	0.08	0.06	0.01	-0.01	0.00	0.00								
	FRA	0.08	0.10	0.07	0.07	0.05	0.06	0.11	0.08	0.01	0.00	0.01	0.01	0.01							
	DEU	0.03	0.05	0.03	0.03	0.02	0.02	0.05	0.04	-0.01	0.00	-0.01	0.00	-0.01	0.00						
	CHE	0.06	0.08	0.04	0.05	0.03	0.04	0.08	0.06	0.01	0.00	0.01	0.02	0.01	0.01	0.00					

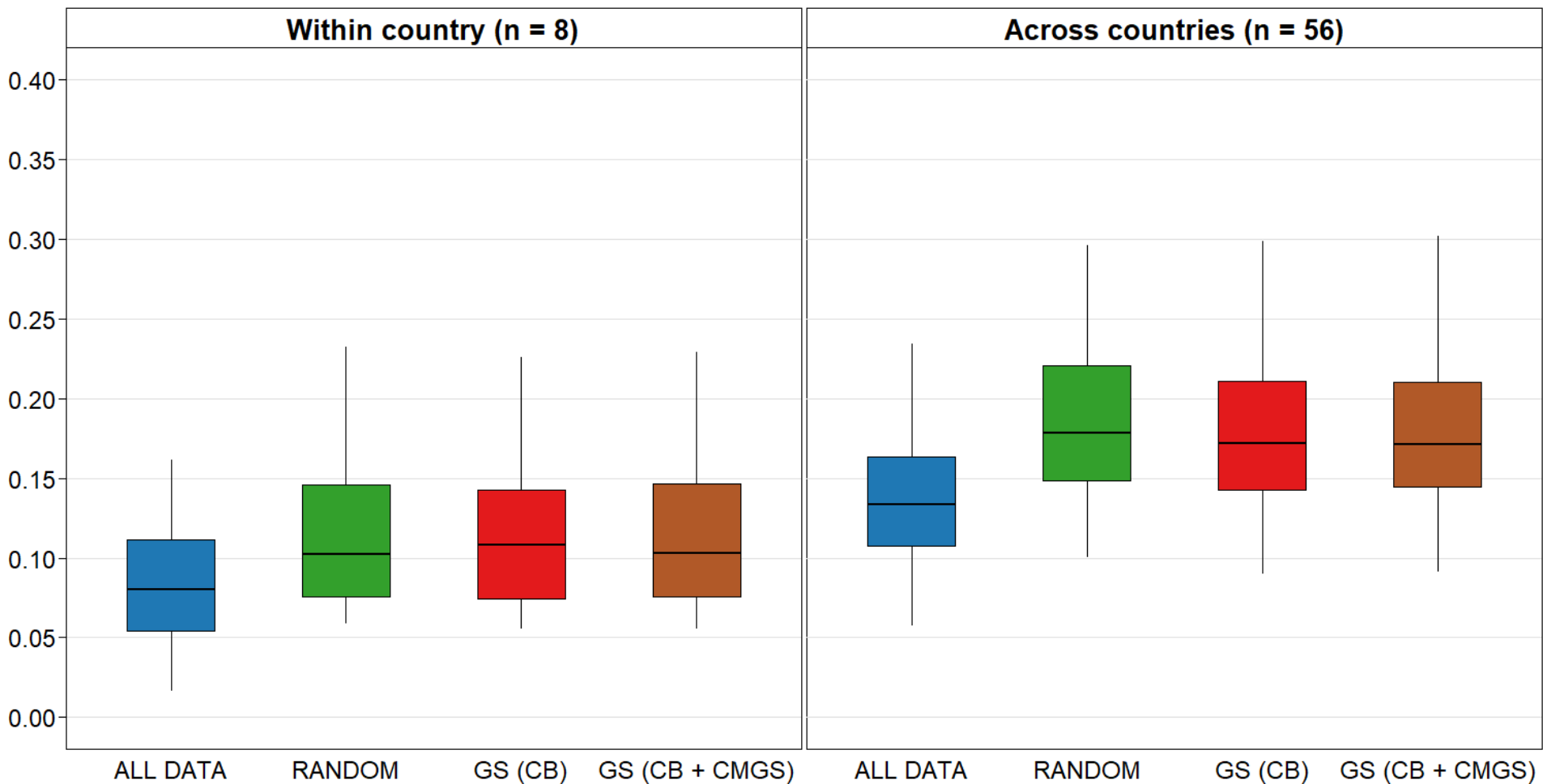
BLUE = non-estimated DMU rG

RED = rG difference (as MIX99 - DMU) > 0.05

Direct and maternal $SE(r_g)$



Direct-maternal $SE(r_g)$



Model

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_8 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{X}_8 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_8 \end{bmatrix} + \begin{bmatrix} \mathbf{C}_1 & 0 & 0 & 0 \\ 0 & \mathbf{C}_2 & 0 & 0 \\ 0 & 0 & \mathbf{C}_3 & 0 \\ 0 & 0 & 0 & \mathbf{C}_4 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \\ \mathbf{r}_4 \end{bmatrix} + \\ \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_8 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_8 \end{bmatrix} + \begin{bmatrix} \mathbf{W}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_8 \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_8 \end{bmatrix} + \begin{bmatrix} \mathbf{P}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{P}_7 \end{bmatrix} \begin{bmatrix} \mathbf{pe}_1 \\ \vdots \\ \mathbf{pe}_7 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_8 \end{bmatrix}$$

$$var \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_8 \\ \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_8 \end{bmatrix} = \begin{bmatrix} \sigma^2_{u1} & & & & & & & & & & \\ \sigma_{u1,u2} & \sigma^2_{u2} & & & & & & & & & \\ \vdots & \vdots & \ddots & & & & & & & & \\ \sigma_{u1,u8} & \sigma_{u1,u8} & \cdots & \sigma^2_{u8} & & & & & & & \\ \sigma_{u1,m1} & \sigma_{u2,m1} & \cdots & \sigma_{u8,m1} & \sigma^2_{m1} & & & & & & \\ \sigma_{u1,m2} & \sigma_{u2,m2} & \cdots & \sigma_{u8,m2} & \sigma_{m1,m2} & \sigma^2_{m2} & & & & & \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & & & & \\ \sigma_{u1,m8} & \sigma_{u2,m8} & \cdots & \sigma_{u8,m8} & \sigma_{m1,m8} & \sigma_{m2,m8} & \cdots & \sigma^2_{m8} & & & \end{bmatrix} \otimes \mathbf{A}$$

$$var(\mathbf{r}) = \bigoplus_{i=1}^4 \mathbf{I}_{ni} \cdot \sigma_{ri}^2, \quad var(\mathbf{pe}) = \bigoplus_{i=1}^7 \mathbf{I}_{ni} \cdot \sigma_{pei}^2, \quad var(\mathbf{e}) = \bigoplus_{i=1}^8 \mathbf{I}_{ni} \cdot \sigma_{ei}^2$$

1. All data genetic correlations (rG) and Standard Errors (SE)

rG		<i>dir-dir</i>		<i>dir-mat (off diag)</i>		SE		<i>dir-dir</i>		<i>dir-mat (off diag)</i>	
average	0.79	average	0.00	average	0.14	average	0.14	average	0.14	average	0.14
min	0.62	min	-0.14	min	0.06	min	0.06	min	0.06	min	0.06
max	0.94	max	0.14	max	0.22	max	0.23	max	0.23	max	0.23

<i>dir-mat (diag)</i>		<i>mat-mat</i>		<i>dir-mat (diag)</i>		<i>mat-mat</i>	
average	-0.12	average	0.71	average	0.09	average	0.19
min	-0.33	min	0.65	min	0.02	min	0.07
max	0.40	max	0.87	max	0.16	max	0.33

Scenario 2. Random selection of FRA herds

rG (average of 3 samples) – rG (all data) and SE

rG		<i>dir-dir</i>		<i>dir-mat (off diag)</i>		SE		<i>dir-dir</i>		<i>dir-mat (off diag)</i>	
average	-0.02	average	-0.11	average	-0.11	average	0.03	average	0.03	average	0.05
min	-0.04	min	-0.16	min	-0.16	min	0.01	min	0.01	min	0.01
max	-0.01	max	-0.06	max	-0.06	max	0.06	max	0.06	max	0.09
<i>dir-mat (diag)</i>		<i>mat-mat</i>		<i>dir-mat (diag)</i>		<i>mat-mat</i>		<i>dir-mat (diag)</i>		<i>mat-mat</i>	
average	-0.12	average	-0.04	average	-0.04	average	0.03	average	0.03	average	0.06
min	-0.17	min	-0.07	min	-0.07	min	0.01	min	0.01	min	0.02
max	-0.04	max	-0.02	max	-0.02	max	0.07	max	0.07	max	0.10

Result (average of 3 runs):

- Overall smaller rG
- *dir-dir* slightly different, *mat-mat* smaller on average
- Biggest differences for *dir-mat* part

Scenario 3. GS at the herd level

rG		<i>dir-dir</i>		<i>dir-mat (off diag)</i>		SE		<i>dir-dir</i>		<i>dir-mat (off diag)</i>	
average	-0.02	average	-0.11	average	-0.11	average	0.03	average	0.03	average	0.04
min	-0.03	min	-0.17	min	-0.17	min	0.00	min	0.00	min	0.01
max	0.00	max	-0.03	max	-0.03	max	0.06	max	0.06	max	0.09
<i>dir-mat (diag)</i>		<i>mat-mat</i>		<i>dir-mat (diag)</i>		<i>mat-mat</i>					
average	-0.13	average	-0.05	average	0.03	average	0.05				
min	-0.17	min	-0.08	min	0.01	min	0.01				
max	-0.04	max	-0.03	max	0.08	max	0.10				

Result (average of 3 runs):

- Overall smaller rG
- *dir-dir* slightly different, *mat-mat* smaller on average
- Biggest differences for *dir-mat* part

Scenario 4. GS at the herd level, with inclusion of Common MGS

rG		<i>dir-dir</i>		<i>dir-mat (off diag)</i>		SE		<i>dir-dir</i>		<i>dir-mat (off diag)</i>	
average	-0.02	average	-0.02	average	-0.12	average	0.02	average	0.02	average	0.04
min	-0.03	min	-0.03	min	-0.17	min	0.00	min	0.00	min	0.01
max	0.00	max	0.00	max	-0.04	max	0.05	max	0.05	max	0.08
<i>dir-mat (diag)</i>		<i>mat-mat</i>		<i>dir-mat (diag)</i>		<i>mat-mat</i>		<i>dir-mat (diag)</i>		<i>mat-mat</i>	
average	-0.13	average	-0.13	average	-0.05	average	0.03	average	0.03	average	0.05
min	-0.18	min	-0.18	min	-0.08	min	0.01	min	0.01	min	0.02
max	-0.05	max	-0.05	max	-0.03	max	0.07	max	0.07	max	0.09

Results: No relevant differences in rG or SE