EDA: MS ComBat GAM vs Linear

Virgilio Gonzenbach

2/13/2021

Objective

To compare effects of applying ComBat-GAM vs ComBat-Linear in the MS data.

IMPORTANT

MS patients from the *HSC* site that were scanned in a *SIEMENSTIMTRIO* scanner were excluded from the current harmonization.

```
[1] "In MS:"
```

	CHP	HSC
SIEMENSPRISMAFIT	0	28
SIEMENSTIMTRIO	0	18
SIEMENSVERIO	21	0

IMPORTANT (cont.)

The corresponding group in the Healthy Control data was similarly excluded when modeling site effects with ComBat-GAM.

```
[1] "In HC:"
```

	CHP	HSC	PNC
SIEMENSPRISMAFIT	0	58	0
SIEMENSTIMTRIO	0	7	1185
SIEMENSVERIO	36	0	0

Datasets: MS

MS (dimensions):

```
## [1]  67 161
```

MS (count by site)

site	n
CHP	21
HSC-SIEMENSPRISMAFIT	28
HSC-SIEMENSTIMTRIO	18

MS (count by sex)

sex	site	n
FEMALE	CHP	18
FEMALE	HSC-SIEMENSPRISMAFIT	17
FEMALE	HSC-SIEMENSTIMTRIO	15
MALE	CHP	3
MALE	HSC-SIEMENSPRISMAFIT	11
MALE	HSC-SIEMENSTIMTRIO	3

Datasets: HC

Dimensions HC (no PNC):

```
## [1] 101 161
```

Count per site:

site	n
CHP	36
HSC-SIEMENSPRISMAFIT	58
HSC-SIEMENSTIMTRIO	7

Count by sex and site:

site	sex	n
CHP	FEMALE	24
CHP	MALE	12
HSC-SIEMENSPRISMAFIT	FEMALE	37
HSC-SIEMENSPRISMAFIT	MALE	21
HSC-SIEMENSTIMTRIO	FEMALE	5
HSC-SIEMENSTIMTRIO	MALE	2

Datasets: HC + MS

Dimensions (no PNC):

```
## [1] 168 161
```

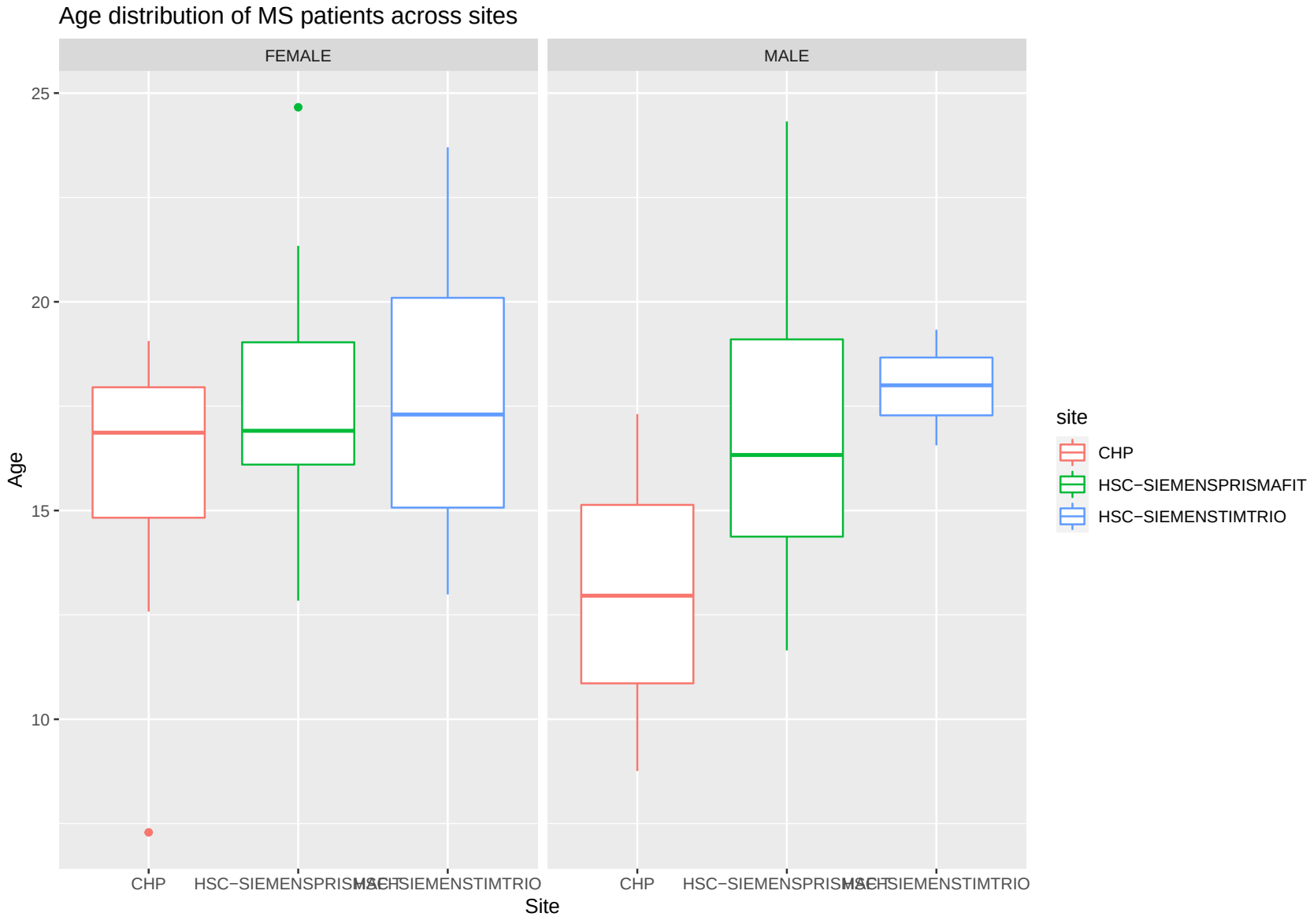
Count per site:

site	n
CHP	57
HSC-SIEMENSPRISMAFIT	86
HSC-SIEMENSTIMTRIO	25

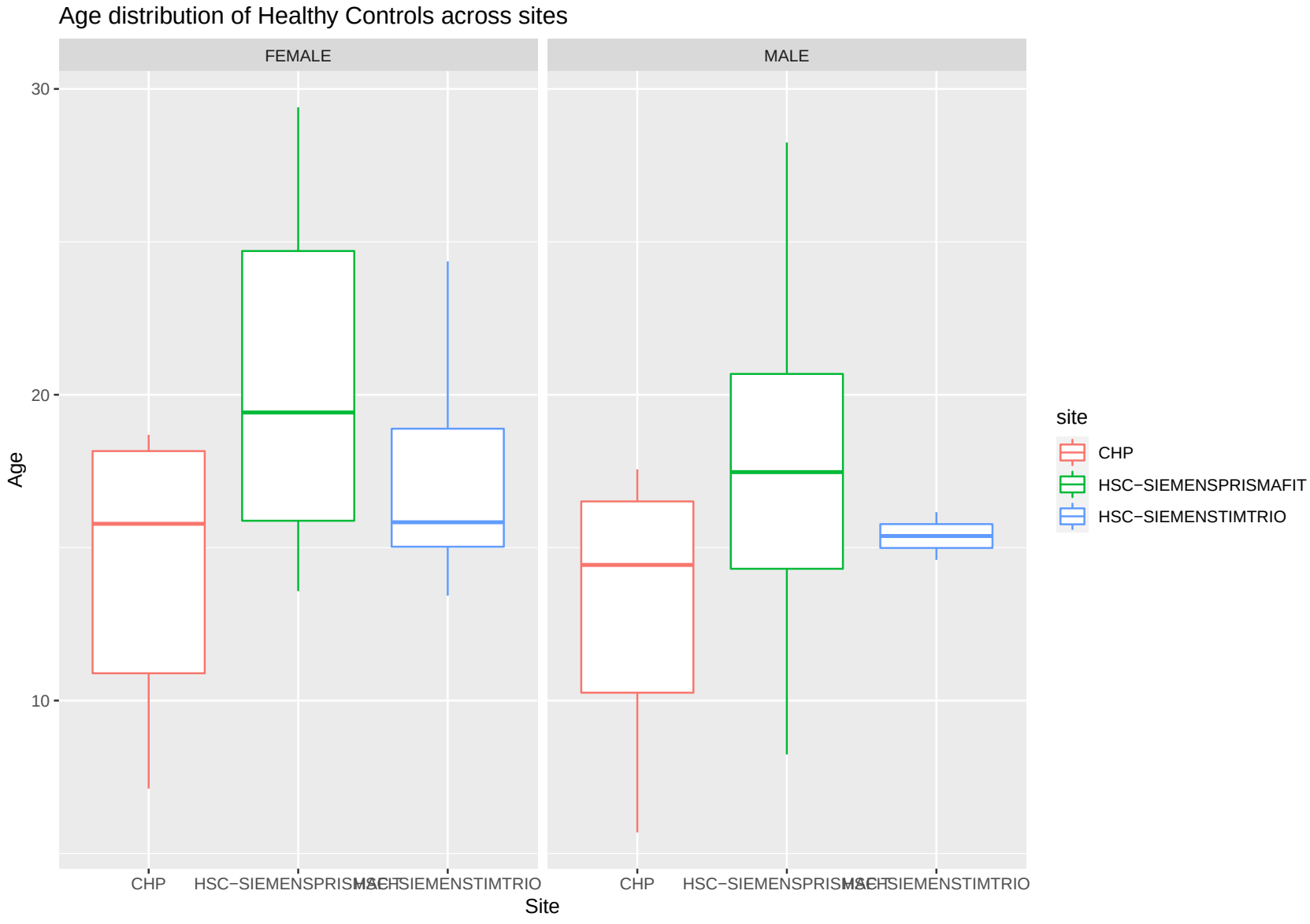
Count by sex and site:

site	sex	n
CHP	FEMALE	42
CHP	MALE	15
HSC-SIEMENSPRISMAFIT	FEMALE	54
HSC-SIEMENSPRISMAFIT	MALE	32
HSC-SIEMENSTIMTRIO	FEMALE	20
HSC-SIEMENSTIMTRIO	MALE	5

Age across sites [MS]



Age across sites [HC]



Harmonization Approach

Adjusted data are shown for the following approach:

- Join HC and MS data into one dataset
- Split this dataset into males and females
- Run ComBat (GAM and Linear) on in parallel on males and females.

Models used in harmonization

- GAM: s(age) + MS + MS*x*age
- Linear: age + age^2 + MS + MS*x*age + MS*x*age2

Site effects: MS + HC [GAM]

Number of ROIs showing site effects:

Full covariate model:

[1] "~ ICV + age + age2 + sex + sex*age* + *sex*age2 + MS + MS*age* + *MS*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Ignore sex: [1] "~ ICV + age + age2 + MS + MS*age* + *MS*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects by sex: MS + HC [GAM]

Females:

[1] "~ ICV + age + age2 + MS + MS*age* + *MS*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Males:

[1] "~ ICV + age + age2 + MS + MS*age* + *MS*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects: MS + HC [Linear]

Full covariate model:

[1] "~ ICV + age + age2 + MS + MS*age* + *MS*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Sex not considered: [1] "~ ICV + age + age2 + MS + MS*age* + *MS*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects by sex: MS + HC [Linear]

Females:

[1] "~ ICV + age + age2 + MS + MS*age* + *MS*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Males:

[1] "~ ICV + age + age2 + MS + MS*age* + *MS*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects: MS [GAM]

Full model:

[1] "~ ICV + age + age2 + sex + sex*age* + *sex*age2"

FDR	Bonferroni	Uncorrected P
0	0	1

Ignore sex:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	1

Site effects: MS [Linear]

Number of ROIs showing site effects:

Full model: [1] "~ ICV + age + age2 + sex + sex*age* + *sex*age2"

FDR	Bonferroni	Uncorrected P
0	0	2

Sex not considered: [1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	2

Site effects by sex: MS [GAM]

Females:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Males:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	8

Site effects by sex: MS [Linear]

Females:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Males:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	8

Site effects: HC [GAM]

Full model:

[1] "~ ICV + age + age2 + sex + sex*age* + *sex*age2"

FDR	Bonferroni	Uncorrected P
0	0	1

Sex not considered:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	1

Site effects by sex: HC [GAM]

Females:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	2

Males:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	1

Site effects: HC [Linear]

Number of ROIs showing site effects:

Full model: [1] "~ ICV + age + age2 + sex + sex*age* + *sex*age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Sex not considered: [1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects by sex: HC [Linear]

Females:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	1

Males:

[1] "~ ICV + age + age2"

FDR	Bonferroni	Uncorrected P
0	0	1

Harmonization Conclusions

Harmonization is effective.

Signal check

$$\text{Full model} = \text{ROI} \sim \text{ICV} + \text{age} + \text{age2} + \text{sex} + \text{sex}\textit{age} + \textit{sex}\text{age2} + \text{MS} + \text{MS}\textit{age} + \textit{MS}\text{age2}$$

Raw data

- **MS**:

FDR	Bonferroni	Uncorrected Ps
0	0	10

- **age:MS**:

FDR	Bonferroni	Uncorrected Ps
0	0	10

- **age2:MS**:

FDR	Bonferroni	Uncorrected Ps
0	0	12

Harmonized data (GAM)

- **MS**:

FDR	Bonferroni	Uncorrected Ps
0	0	10

- **age:MS**:

FDR	Bonferroni	Uncorrected Ps
0	0	10

- **age2:MS**:

FDR	Bonferroni	Uncorrected Ps
0	0	12

Raw data (w/ site)

- **MS**:

FDR	Bonferroni	Uncorrected Ps
0	0	9

- **age:sexMALE**:

FDR	Bonferroni	Uncorrected Ps
0	0	18

- **age2:sexMALE**:

FDR	Bonferroni	Uncorrected Ps
0	0	4

Harmonized data (w/ site)

- **MS**:

FDR	Bonferroni	Uncorrected Ps
0	0	9

- **age:sexMALE**:

FDR	Bonferroni	Uncorrected Ps
0	0	18

- **age2:sexMALE**:

FDR	Bonferroni	Uncorrected Ps
0	0	4

Signal check conclusion

Harmonization does not change the number of significant p-values when testing the full model. Success!