MSKIDS Scanner Harmonization

Virgilio Gonzenbach, Kelly Clark, Taki Shinohara

04/09/2021

Overview

- MSKIDS data summary.
- Finding from PNC analysis guiding harmonization approach.
- Harmonization of scanner effects with ComBat-GAM and post-harmonization evaluation.

Section 1

MSKIDS data summary

MSKIDS: All participants

Per scanner:

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

Per sex and site:

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

Totals: Females = 116; Males = 52; All = 168

MSKIDS: HC only

Per site:

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

Per sex and site:

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

Totals: Females = 66; Males = 35: All = 101

MSKIDS: MS only

Per site:

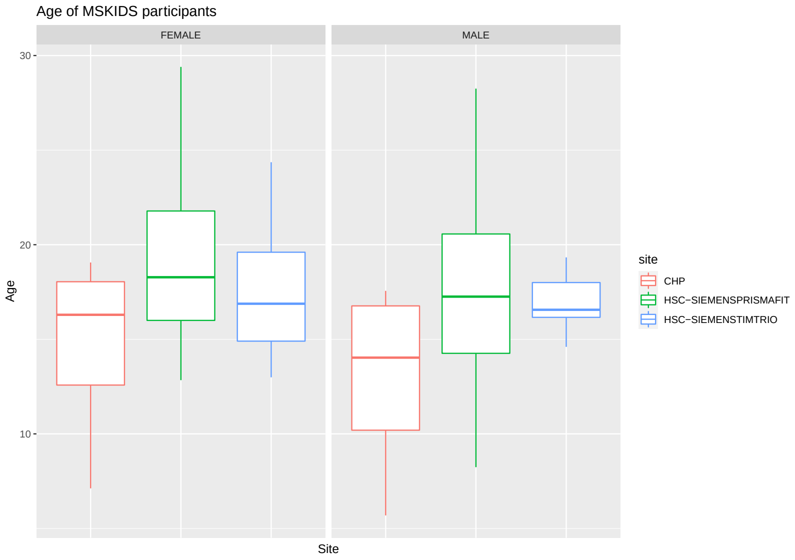
site	n
CHP	21
HSC-SIEMENSPRISMAFIT	28
HSC-SIEMENSTIMTRIO	18

Per sex and site:

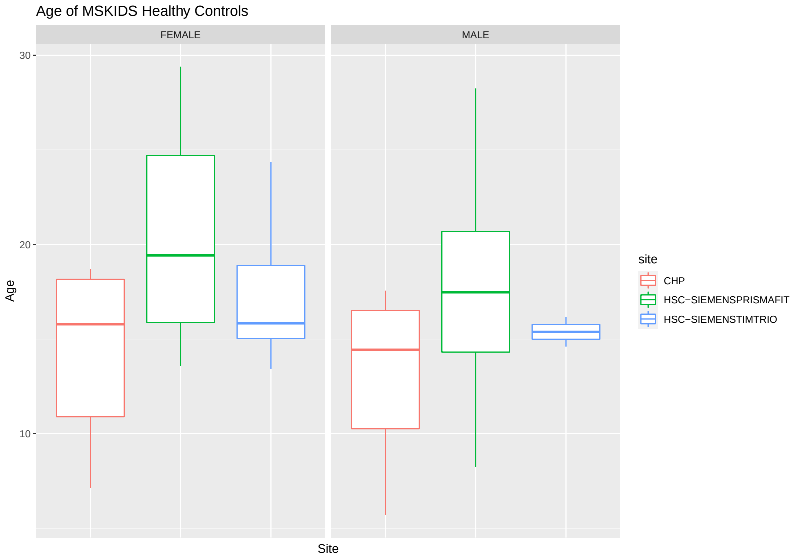
sex	site	n
FEMALE	CHP	18
FEMALE	HSC-SIEMENSPRISMAFIT	17
FEMALE	HSC-SIEMENSTIMTRIO	15
MALE	CHP	3
MALE	HSC-SIEMENSPRISMAFIT	11
MALE	HSC-SIEMENSTIMTRIO	3

Totals: Females = 50; Males = 17; All = 67

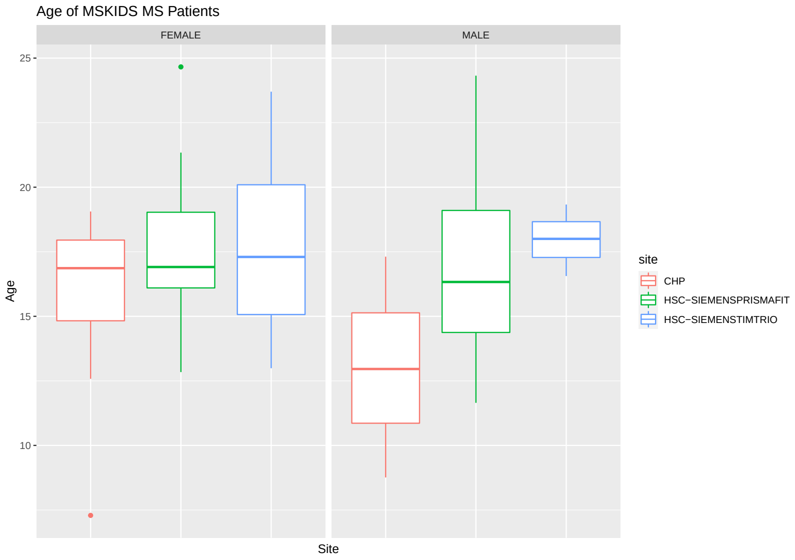
MSKIDS: Age [All participants]



MSKIDS: Age [HC only]



MSKIDS: Age [MS only]



Section 2

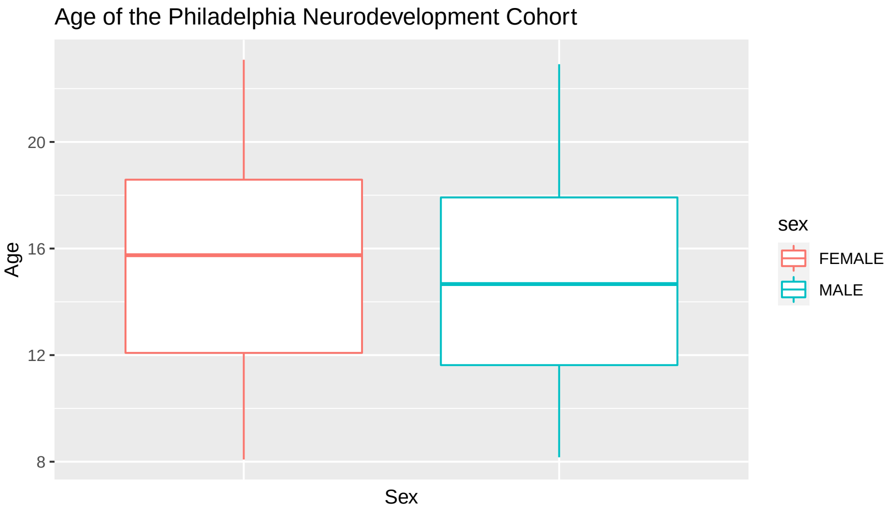
PNC Analysis

Overview

Objective: To inform harmonization approach by

- Determining Age-ROI relationships in Males and Females
- Ascertaining non-linear age trends in select ROIs

Age by Sex



Females: 630 Males : 555

Age descriptives

Full dataset:

	n	mean	sd	median	min	max
age	1185	15.12	3.742	15.33	8.083	23.08

Males:

	n	mean	sd	median	min	max
age	555	14.78	3.717	14.67	8.167	22.92

Females:

	n	mean	sd	median	min	max
age	630	15.42	3.741	15.75	8.083	23.08

ICV: Models by Sex

Table 4

	Dependent variable:	
	Intracranial Volume	
	Males	Females
	(1)	(2)
Age	0.118** (0.035, 0.201)	−0.037 (−0.116, 0.041)
Constant	−0.000 (−0.083, 0.083)	0.000 (−0.078, 0.078)
R^2	0.014	0.001
Note:		*p<0.05; **p<0.01; ***p<0.001

White Matter: Models by Sex

Table 5

	Dependent variable:	
	White Matter Volume	
	Males	Females
	(1)	(2)
Age	0.467^{***} (0.393, 0.541)	0.374^{***} (0.302, 0.447)
Constant	-0.000 (-0.074, 0.074)	-0.000 (-0.072, 0.072)
R^2	0.218	0.140
Note:		$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Gray Matter: Models by Sex

Table 6

	Dependent variable:	
	Gray Matter Volume	
	Males	Females
	(1)	(2)
Age	-0.245^{***} (-0.326, -0.164)	-0.402^{***} (-0.473, -0.330)
Constant	0.000 (-0.081, 0.081)	-0.000 (-0.072, 0.072)
R^2	0.060	0.161
Note:		$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

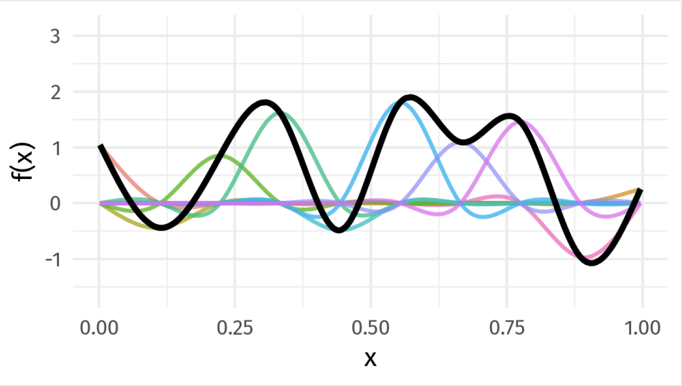
CSF (in ventricles): Models by Sex

Table 7

	Dependent variable:	
	Cerebrospinal Fluid Volume	
	Males	Females
	(1)	(2)
Age	0.284^{***} (0.204, 0.364)	0.171^{***} (0.094, 0.248)
Constant	−0.000 (−0.080, 0.080)	0.000 (−0.077, 0.077)
R^2	0.081	0.029
Note:		$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

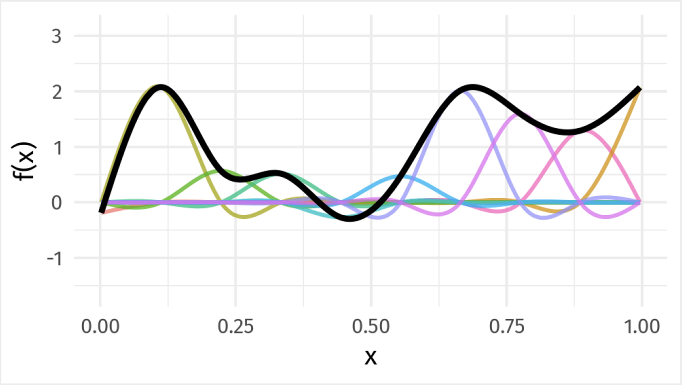
Analysis of non-linear age trends: Generalized Additive Models (1)

Generalized Additive Models (GAMs) allow for modeling non-linear relationships through the use of smooth functions composed of adjustable basis functions.



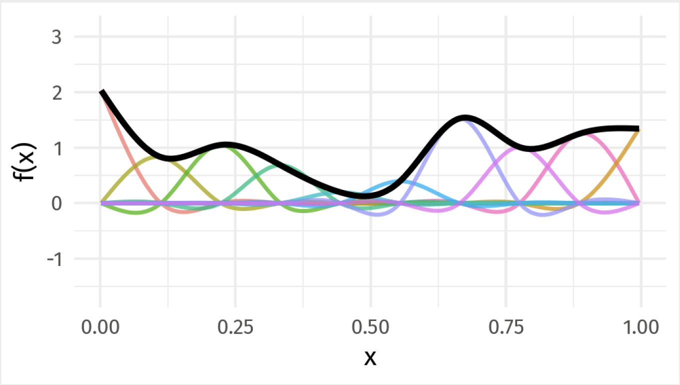
Analysis of non-linear age trends: Generalized Additive Models (2)

Generalized Additive Models (GAMs) allow for modeling non-linear relationships through the use of smooth functions composed of adjustable basis functions.



Analysis of non-linear age trends: Generalized Additive Models (3)

Generalized Additive Models (GAMs) allow for modeling non-linear relationships through the use of smooth functions composed of adjustable basis functions.



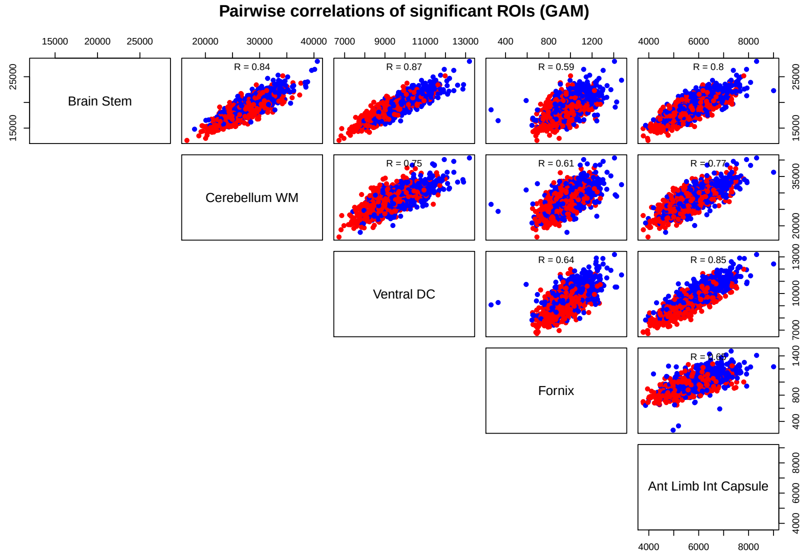
Comparison of GAMs vs Linear models across all 145 ROIs

- Model 1: ROI = sex + age + age*sex
- Model 2: ROI = sex + s(age, by = sex), where s() denotes the smooth function(s) fitted separately within each sex.

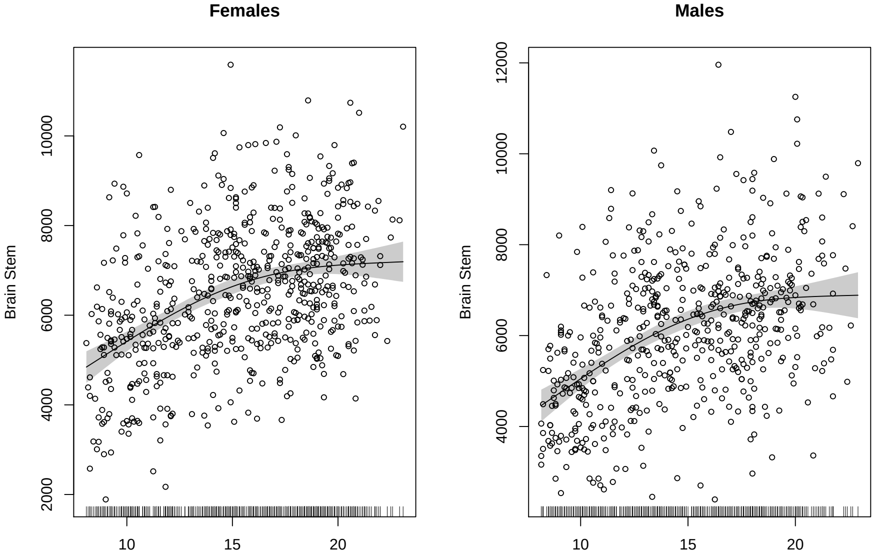
ROIs showing non-linear age effects (FDR-adjusted)

	ROI Name	Hemisphere	Tissue
35	Brain Stem	Both	NONE
40	Right Cerebellum White Matter	Right	WM
41	Left Cerebellum White Matter	Left	WM
61	Right Ventral DC	Right	WM
62	Left Ventral DC	Left	WM
89	fornix right	Right	WM
90	fornix left	Left	WM
91	anterior limb of internal capsule right	Right	WM
92	anterior limb of internal capsule left	Left	WM

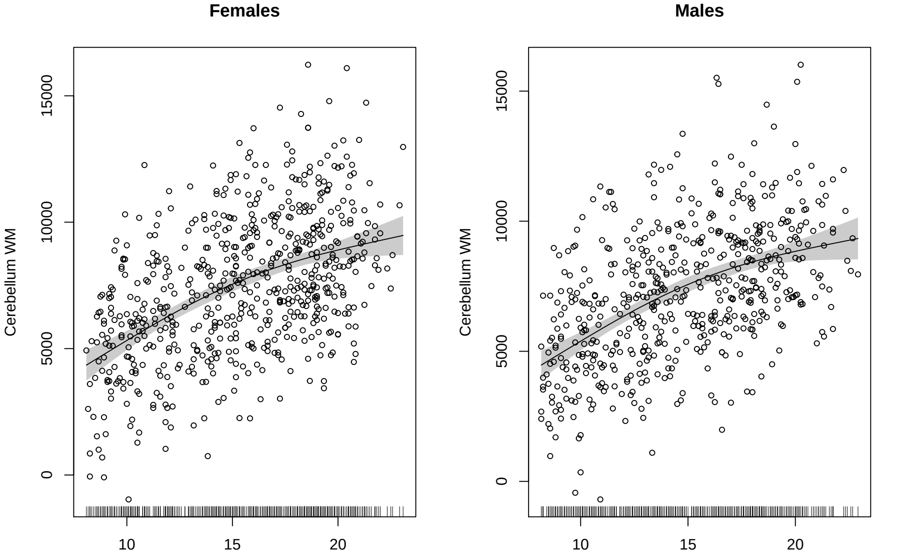
Pairwise Correlations



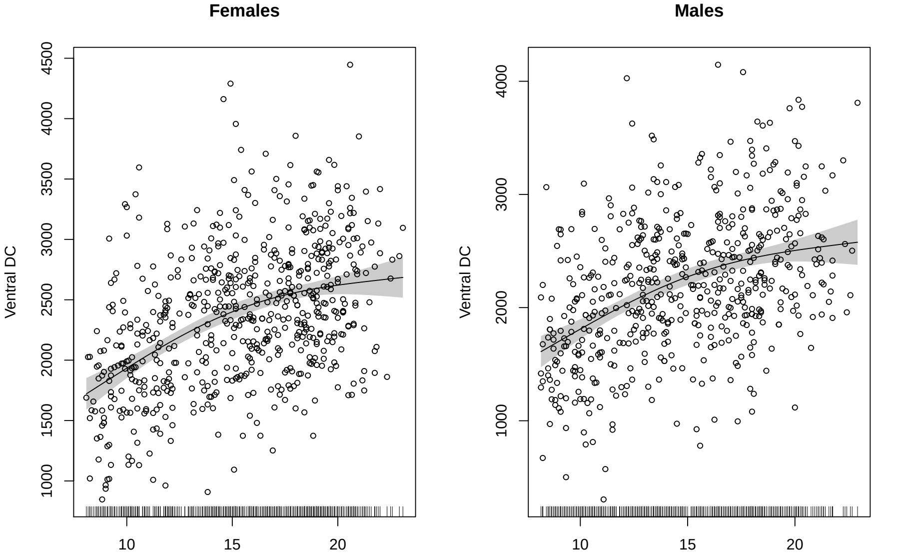
Brain Stem: GAM Plot



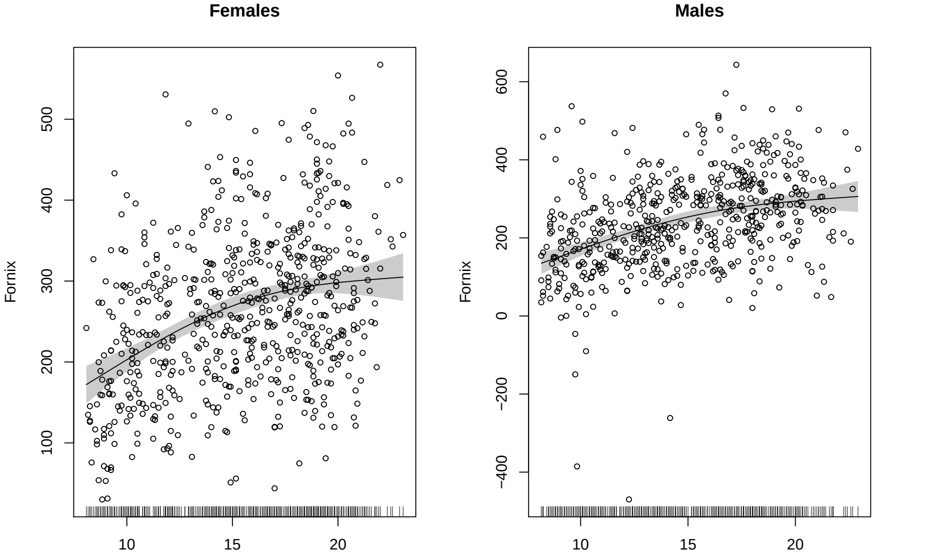
Cerebellum WM: GAM plot



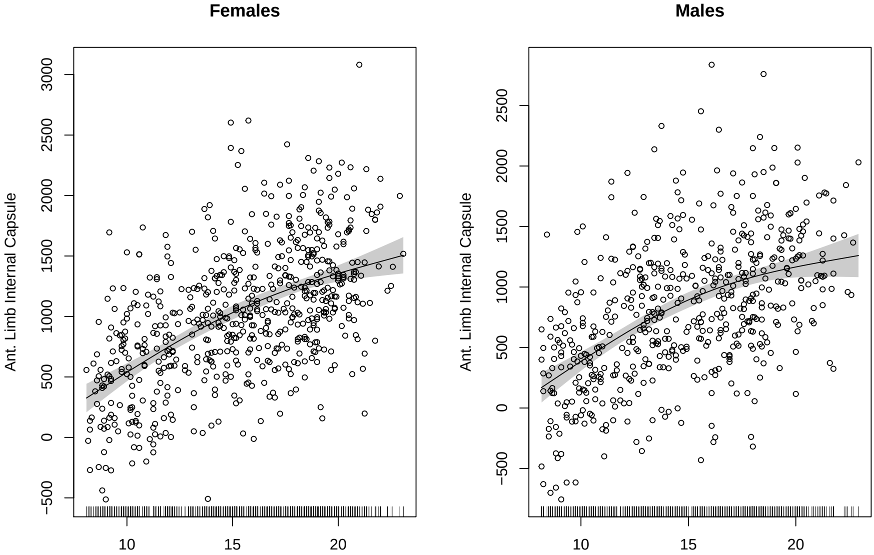
Ventral Diencephalon: GAM Plot



Fornix: GAM Plot



Anterior Limb of Internal Capsule (ALIC): GAM Plot



Section 3

Harmonization

Factors guiding harmonization approach

- Differential Age confound in Males vs. Females.
- ICV, ROI volume differences in Males vs. Females.
- Differential ROI-age relationships in Males vs. Females.
- Non-linear age trends in ROIs.

Harmonization Approach

Adjusted data are shown for the following approach:

1. Split dataset according to sex.
2. Harmonize (Step 1): Run ComBat-GAM on ICV.
3. Harmonize (Step 2): Add harmonized ICV as a covariate. Run ComBat-GAM on 145 ROIs.

Model used in harmonization (separately on each sex)

- GAM: ICV + s(age) + MS + MS$x$age

Testing for site effects

ANOVAs were run on each of the 145 ROIs comparing two versions of covariate model: with and without the inclusion of a site/scanner variable.

Site effects: MS + HC

Number of ROIs showing site effects:

Raw data:

[1] "~ ICV + MS + MS*age + s(age, k=3, bs='tp', fx=TRUE, by=sex)"

FDR	Bonferroni	Uncorrected P
41	18	62

Harmonized data:

[1] "~ ICV + MS + MS*age + s(age, k=3, bs='tp', fx=TRUE, by=sex)"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects: MS + HC [Females]

Raw:

[1] "~ ICV + MS + MS*age + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
29	15	47

Harmonized:

[1] "~ ICV + MS + MS*age + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects: MS + HC [Males]

Raw:

[1] "$\sim$ ICV + MS + MS*age + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
15	2	36

Harmonized: [1] "$\sim$ ICV + MS + MS*age + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects: MS

Raw:

[1] "~ ICV + s(age, k=3, bs='tp', fx=TRUE, by=sex)"

FDR	Bonferroni	Uncorrected P
26	11	38

Harmonized: [1] "~ ICV + s(age, k=3, bs='tp', fx=TRUE, by=sex)"

FDR	Bonferroni	Uncorrected P
0	0	5

Site effects: MS [Females]

Raw: [1] "~ ICV + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
3	2	30

Harmonized:

[1] "~ ICV + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
0	0	1

Site effects: MS [Males]

Raw:
[1] "$\sim$ ICV + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
0	0	24

Harmonized:
[1] "$\sim$ ICV + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
0	0	24

Site effects: HC

Raw:

[1] "~ ICV + s(age, k=3, bs='tp', fx=TRUE, by=sex)"

FDR	Bonferroni	Uncorrected P
23	10	37

Harmonized:

[1] "~ ICV + s(age, k=3, bs='tp', fx=TRUE, by=sex)"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effect: HC [Females]

Raw

[1] "$\sim$ ICV + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
15	7	35

Harmonized:

[1] "$\sim$ ICV + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
0	0	0

Site effects: HC [Males]

Raw: [1] "~ ICV + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
1	1	28

Harmonized: [1] "~ ICV + s(age, k=3, bs='tp', fx=TRUE)"

FDR	Bonferroni	Uncorrected P
0	0	0

Harmonization Conclusions

The current implementation of ComBat-GAM is successful at removing site effects in all subsets of the data!!