

Simon Says Stats

“First we take Manhattan, then we take Berlin”

Adjusting for site-effects

Simon White

Department of Psychiatry, University of Cambridge
MRC Biostatistics Unit, University of Cambridge

2020/Jul/14
Ed Group Meeting

Site?

Example 1

Identifiability

Example 2

Example 3

Example 4

Removing site

Summary

● ○○ ○ ○ ○ ○○ ○

What are we trying to do?

What are we trying to do?

Is “site” of interest?

The magnitude of the site-effect is important, relative to?

Is “site” a nuisance?

A confounder of no scientific interest, to be “removed” from our effect estimates somehow?

What are we trying to do?

Is “site” of interest?

The magnitude of the site-effect is important, relative to?

Is “site” a nuisance?

A confounder of no scientific interest, to be “removed” from our effect estimates somehow?

Always more questions...

What are we trying to do?

Is “site” of interest?

The magnitude of the site-effect is important, relative to?

Is “site” a nuisance?

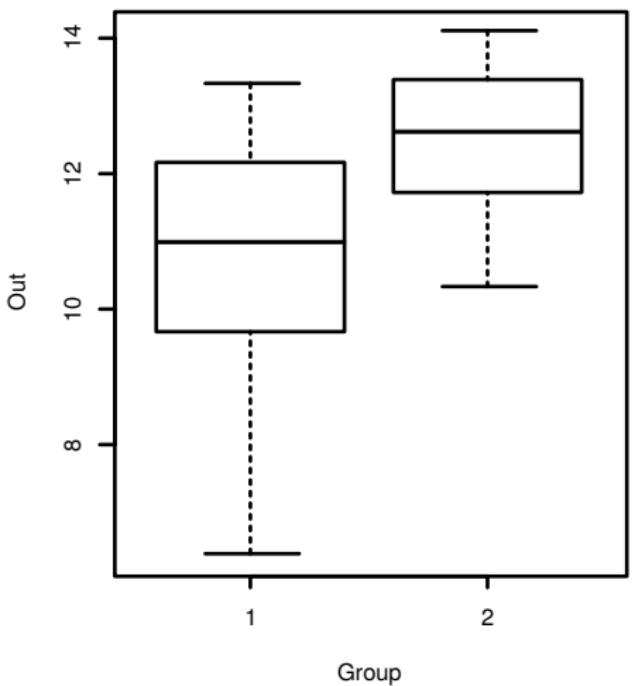
A confounder of no scientific interest, to be “removed” from our effect estimates somehow?

Always more questions...

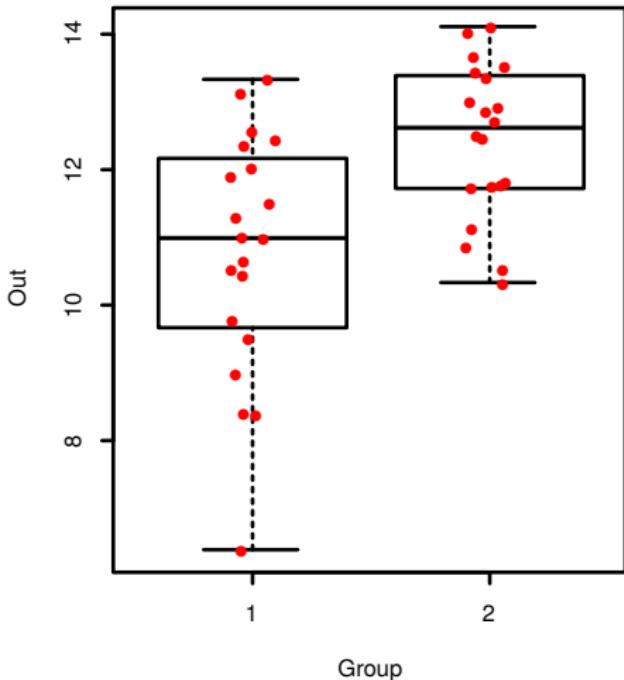
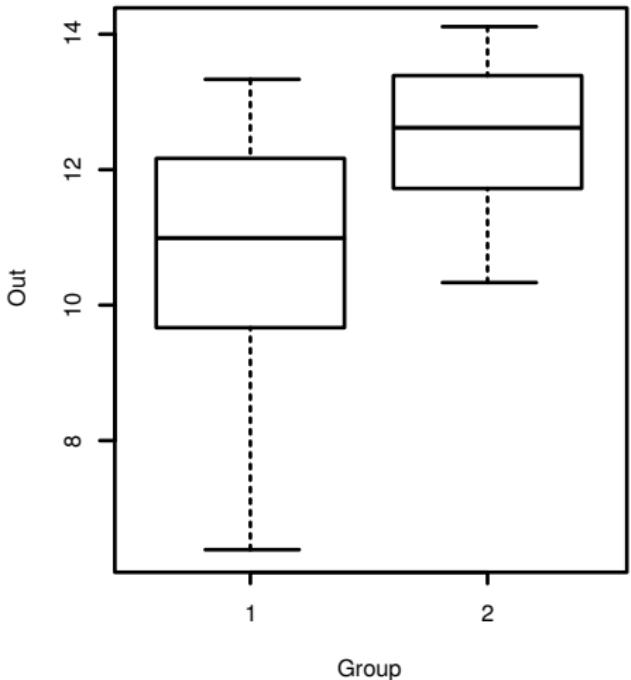
What does “site” mean?

- A directly interpretable concept
- A collection of effects

Simple: two groups with single continuous outcome

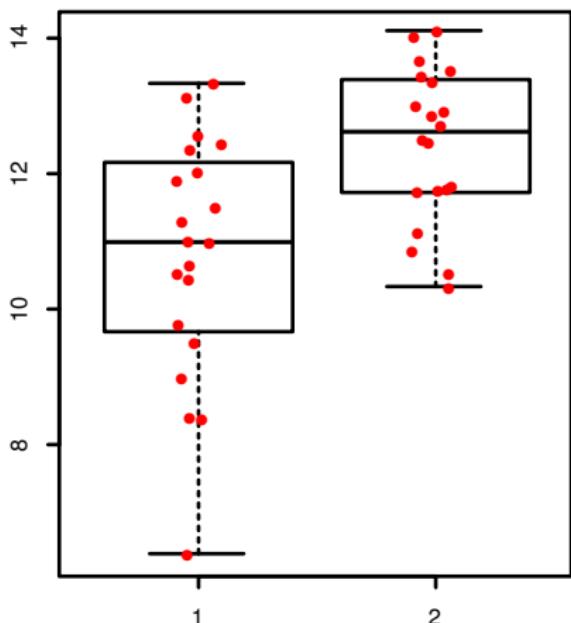


Simple: two groups with single continuous outcome



Simple: two groups with single continuous outcome

$\hat{\mu}_1 = 10.79$
 $\hat{\mu}_2 = 12.42$
diff = 1.638
s.e. = 0.472
df = 38
CI = (0.682, 2.594)
p-value = 0.0013



Site?



Example 1



Identifiability



Example 2



Example 3



Example 4

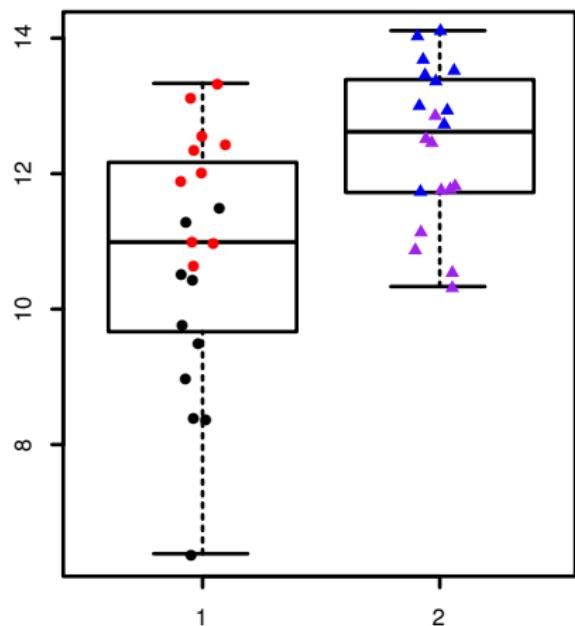


Removing site



Summary

But wait...



Site?



Example 1



Identifiability



Example 2



Example 3



Example 4

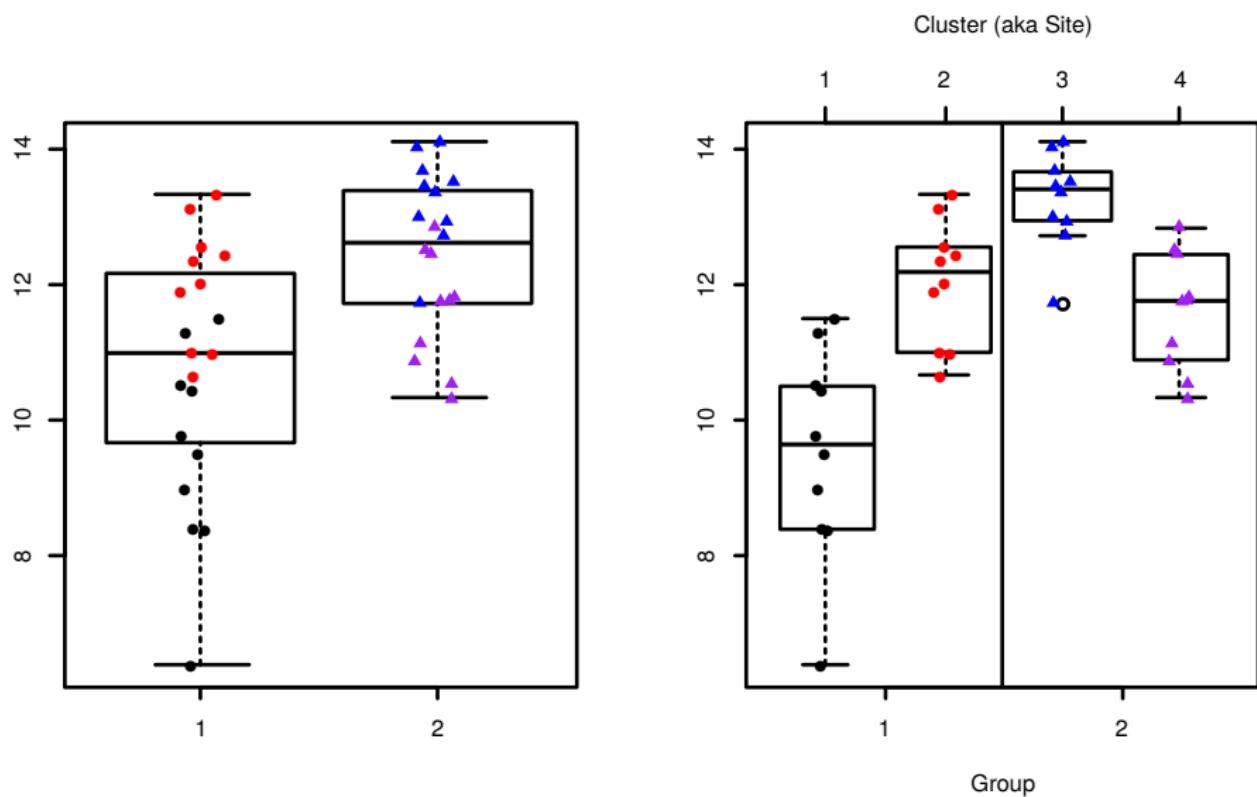


Removing site



Summary

But wait...



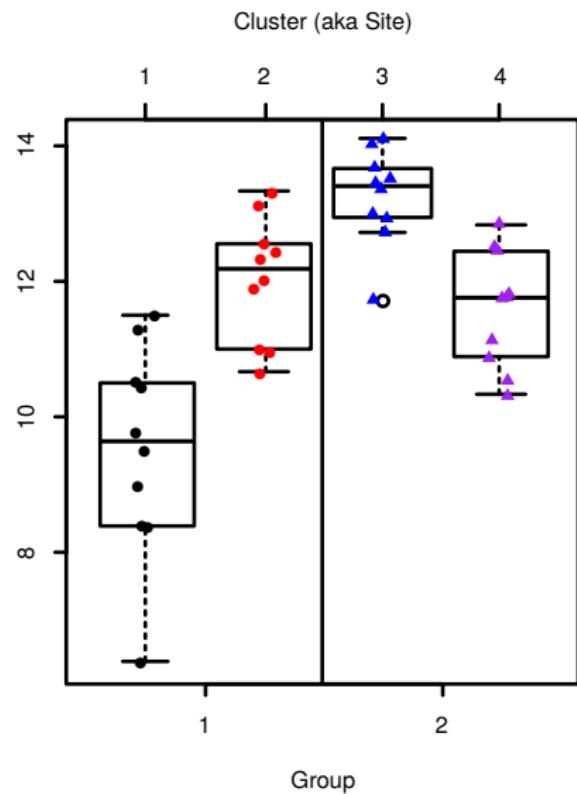
But wait...

Within-cluster dependency

⇒ Intra-class Correlation Coefficient (ICC)

$$\sigma_\alpha = 1.458, \sigma_\epsilon = 1.054$$

$$\text{ICC} = \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + \sigma_\alpha^2} = 0.657$$



But wait...

Within-cluster dependency

⇒ Intra-class Correlation Coefficient (ICC)

$$\sigma_\alpha = 1.458, \sigma_\epsilon = 1.054$$

$$ICC = \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + \sigma_\alpha^2} = 0.657$$

Design Effect (D_{eff})

$$D_{eff} = 1 + ICC \times (m-1)$$

$$D_{eff} = 1 + 0.657(10-1) = 6.913$$

$$diff = 1.638$$

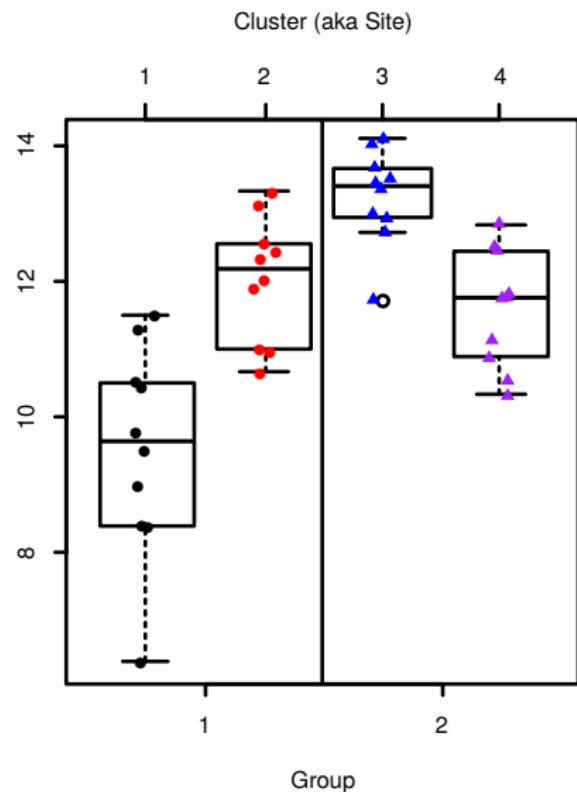
$$s.e. = 0.472$$

$$s.e._{mod} = D_{eff} \times s.e. = 3.264$$

$$df = 38$$

$$CI_{mod} = (-4.97, 8.246)$$

$$p\text{-value} = 0.6188$$

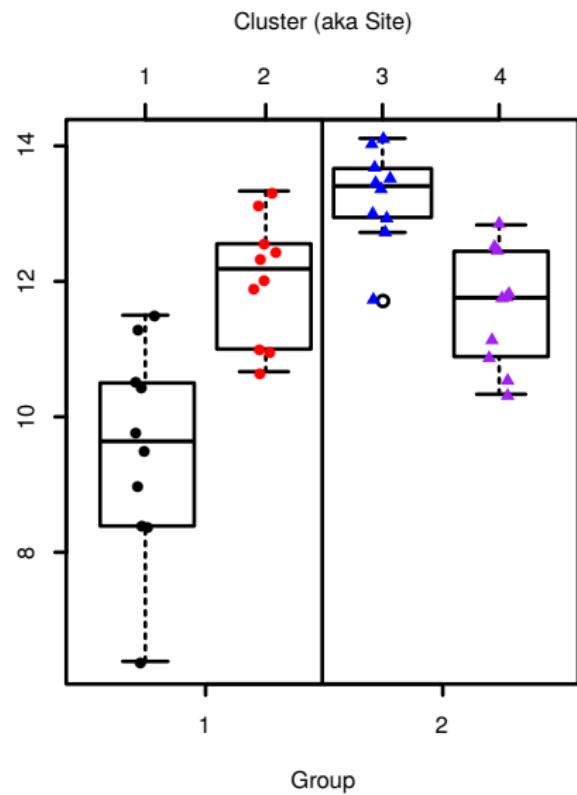


But wait...

Cluster/site-effect can reduce effect sample size

Many ICC variants

Many Deff variants



Site?



Example 1



Identifiability



Example 2



Example 3



Example 4



Removing site



Summary

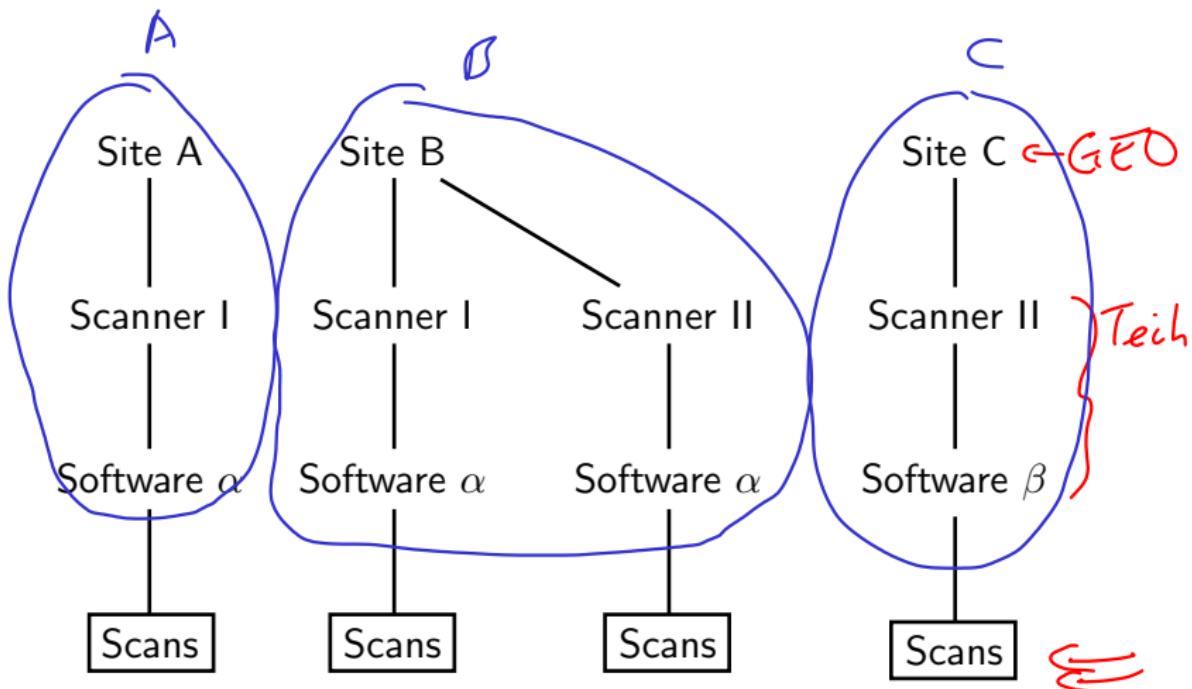


Identifiability

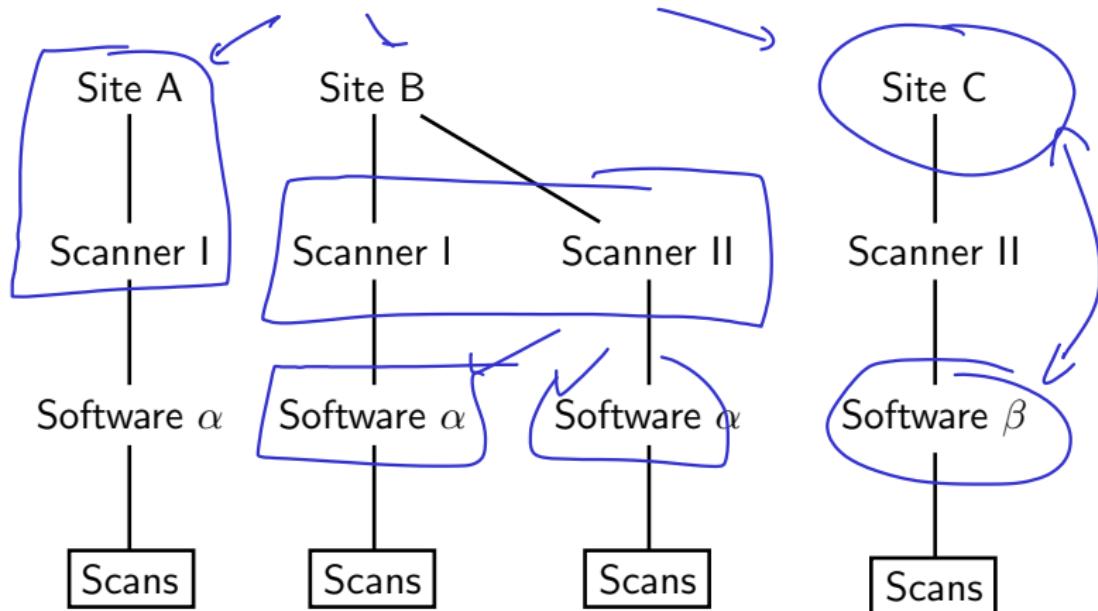
Very easy to define a non-identifiable model

NESTING of levels/factors/covariates is key

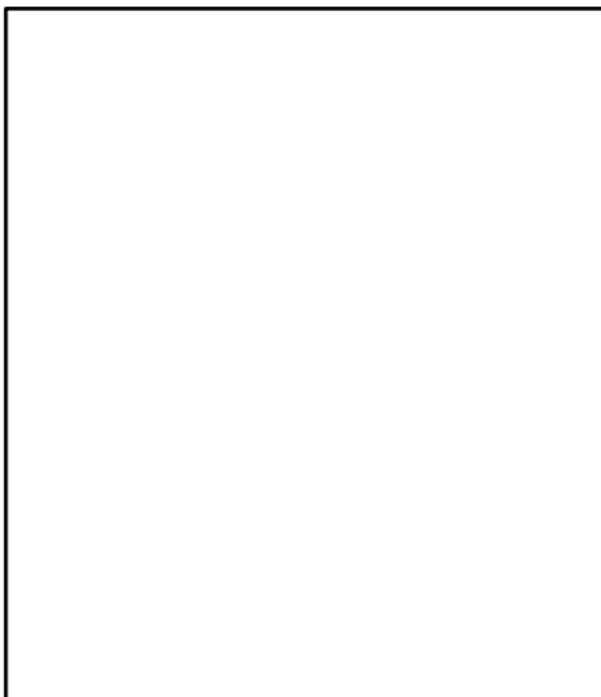
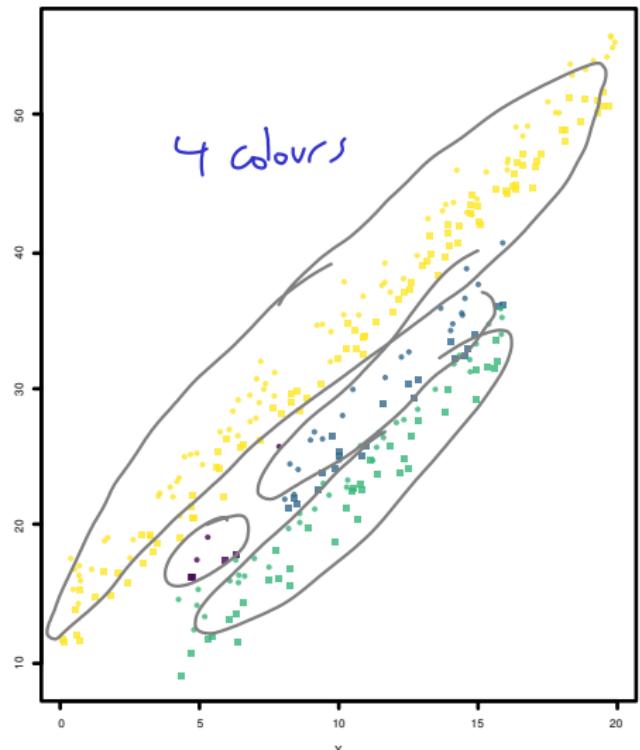
What can we hope to infer?



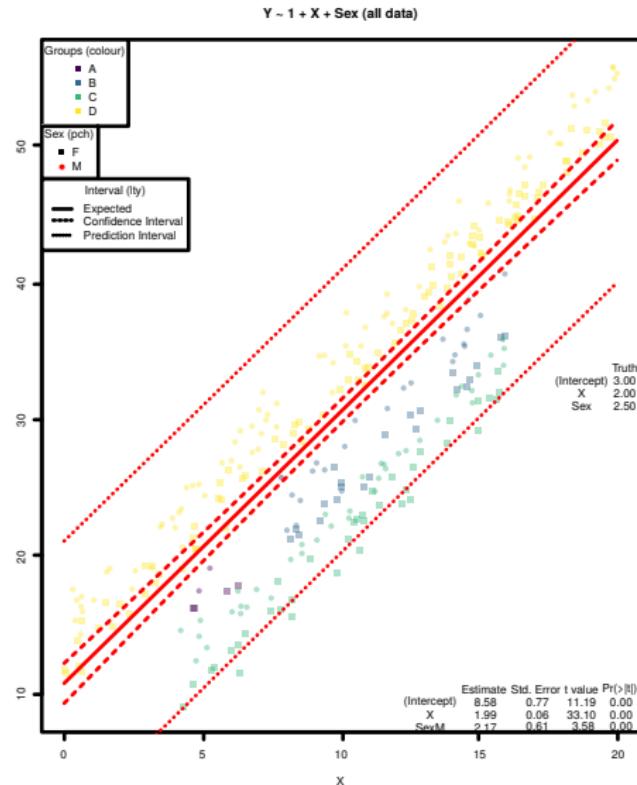
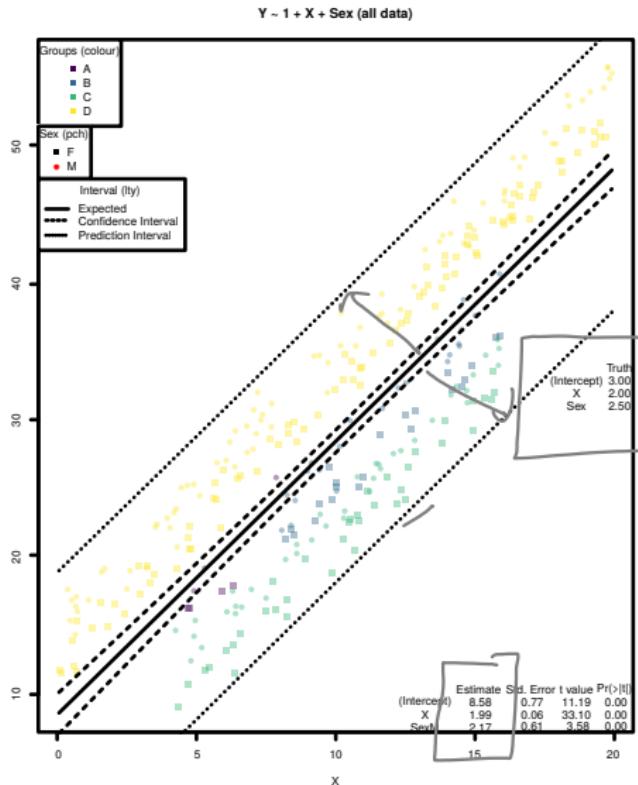
What can we hope to infer?



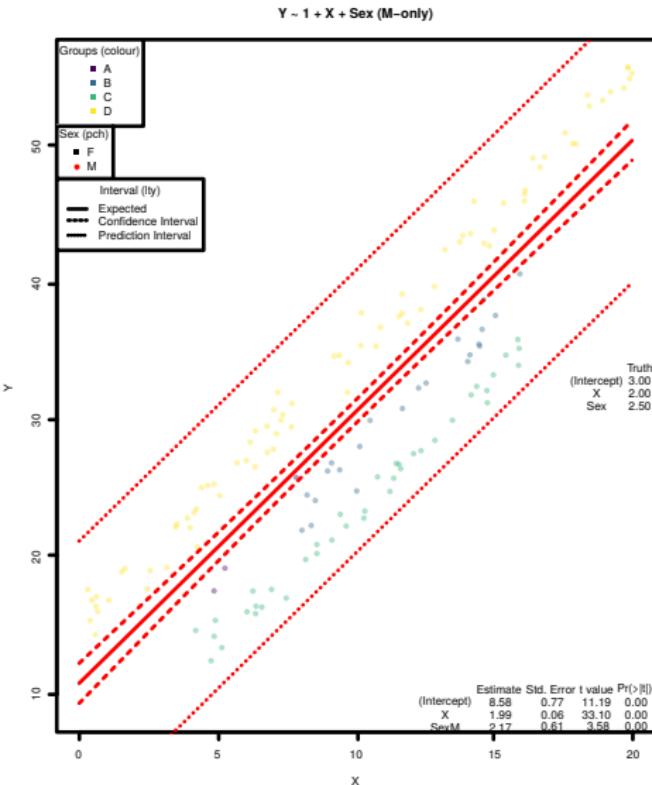
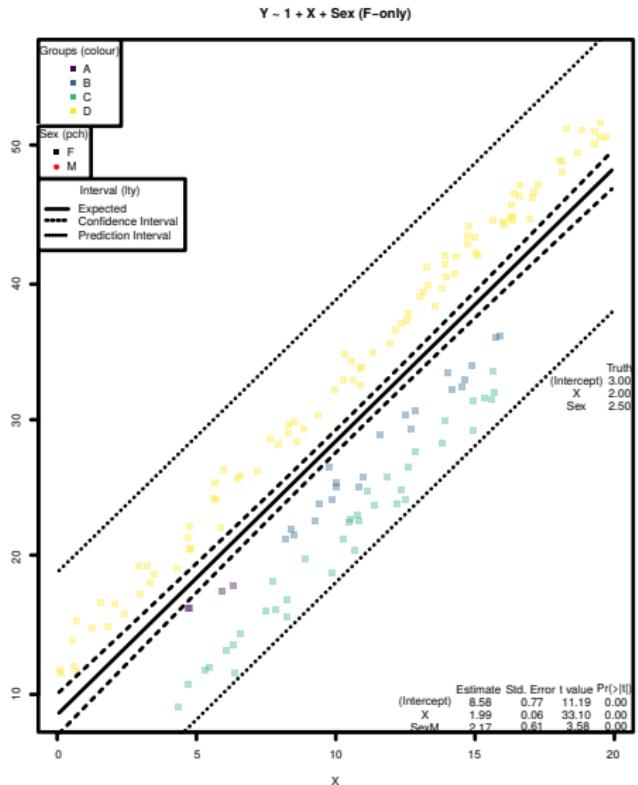
Accounting for multiple studies



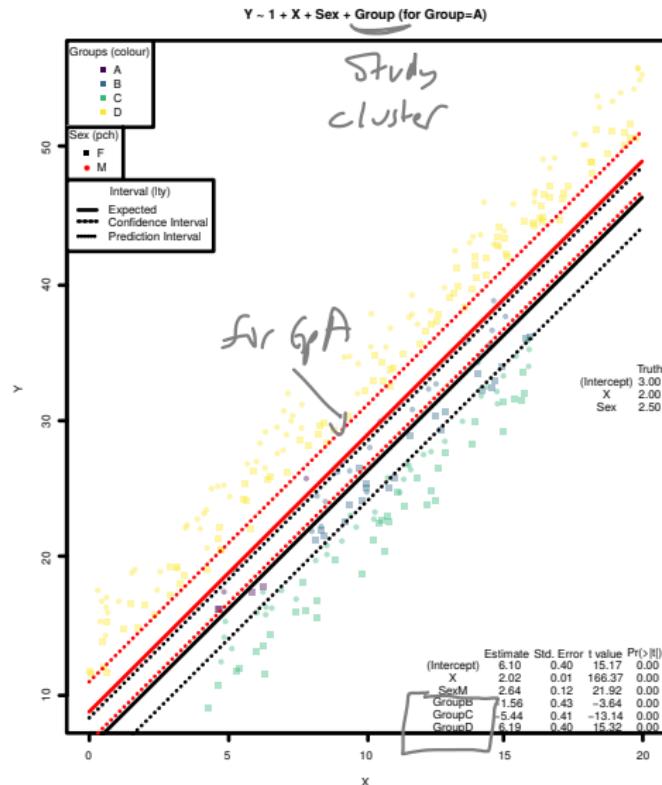
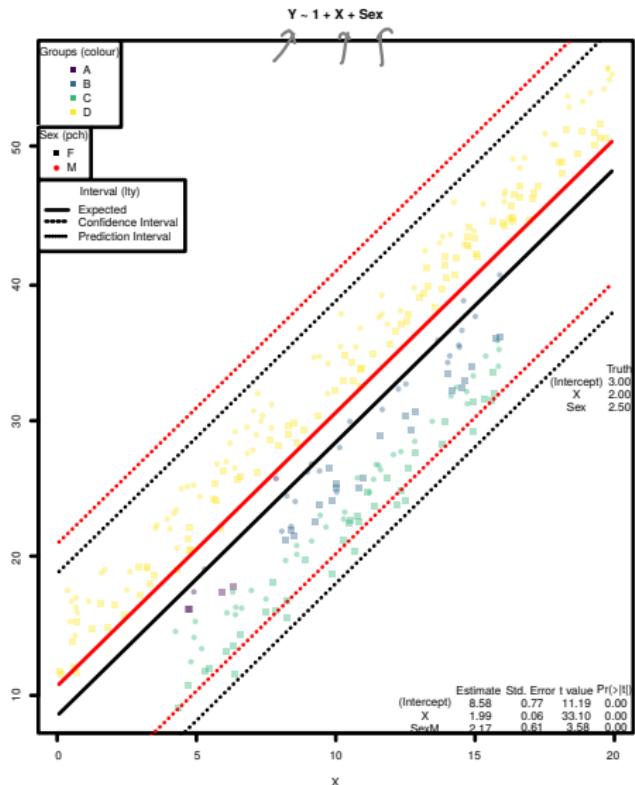
Accounting for multiple studies



Accounting for multiple studies

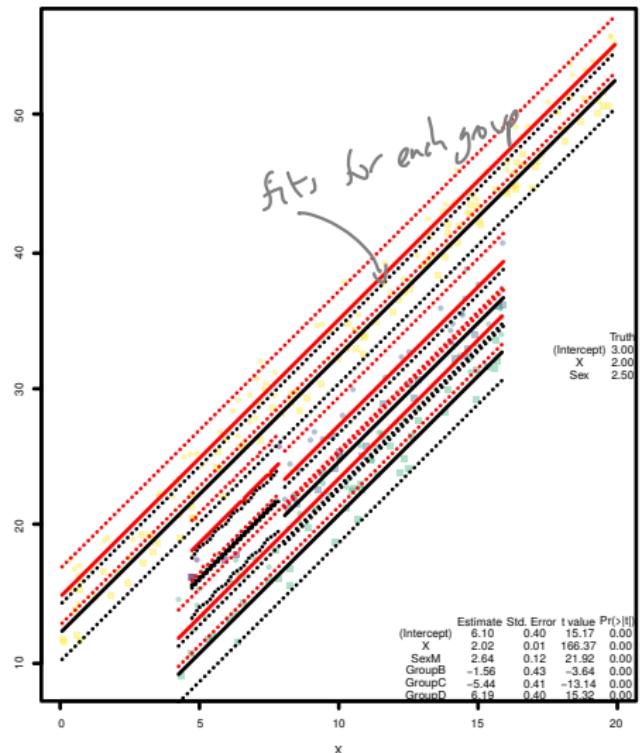


Accounting for multiple studies

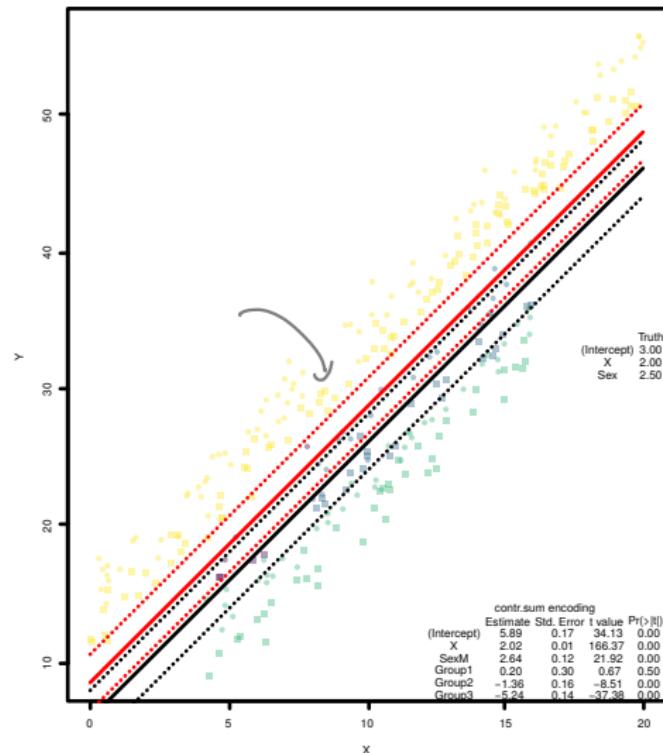


Accounting for multiple studies

$Y \sim 1 + X + \text{Sex} + \text{Group}$ (Group-specific fits)

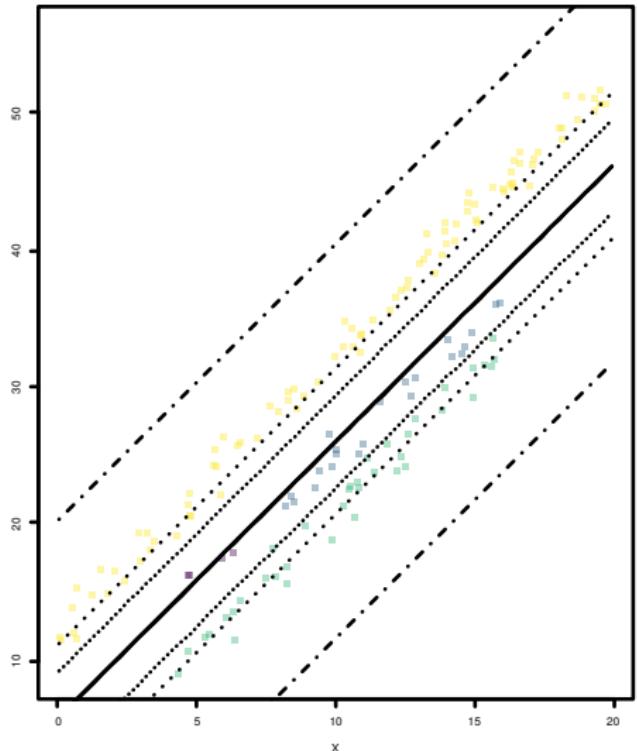


$Y \sim 1 + X + \text{Sex} + \text{Group}$ (Group-average fit)

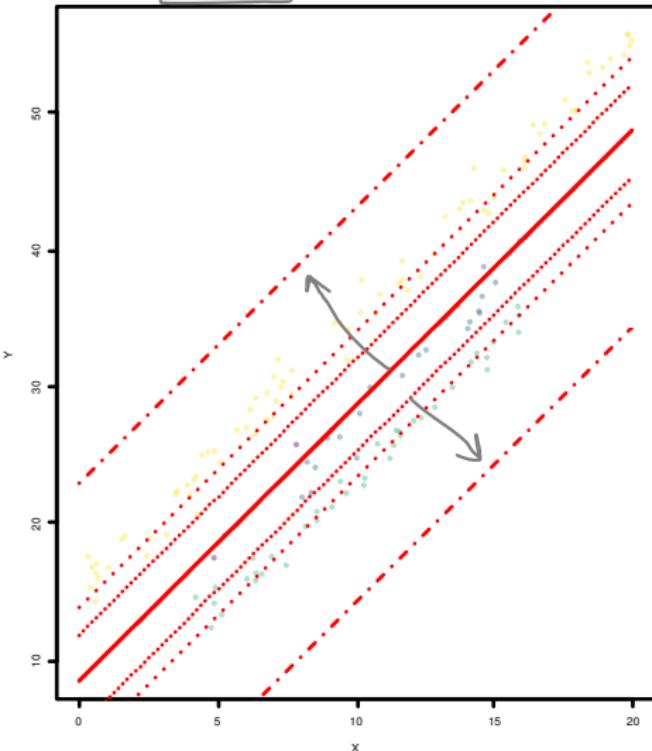


Accounting for multiple studies

$Y \sim 1 + X + Sex + (1|Group)$ [for new/unknown Group]

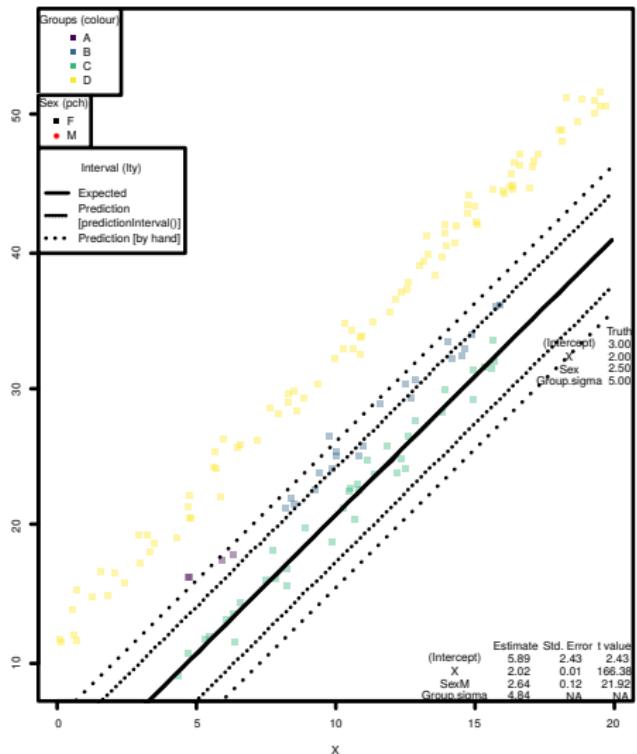


$Y \sim 1 + X + Sex + (1|Group)$ [for new/unknown Group]

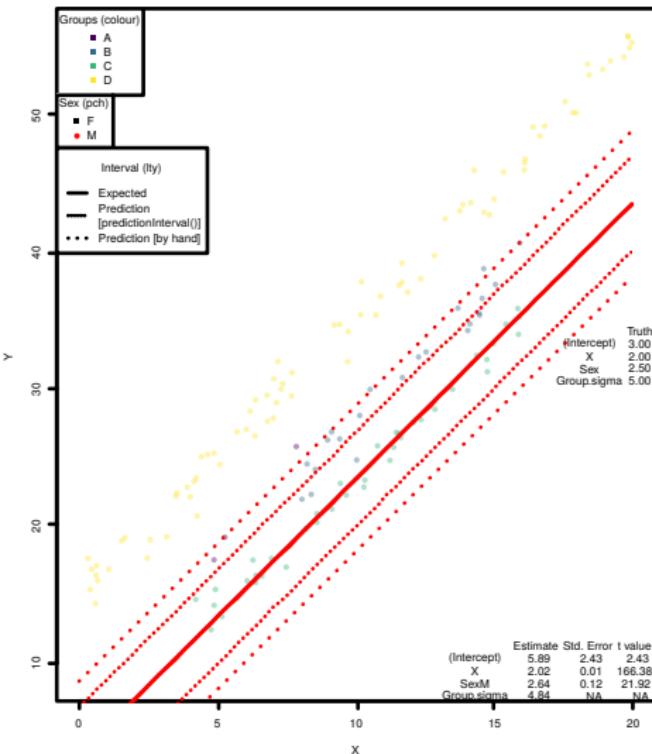


Accounting for multiple studies

$Y \sim 1 + X + Sex + (1|Group)$ [for Known Group=C]



$Y \sim 1 + X + Sex + (1|Group)$ [for Known Group=C]



Site?



Example 1



Identifiability



Example 2



Example 3



Example 4

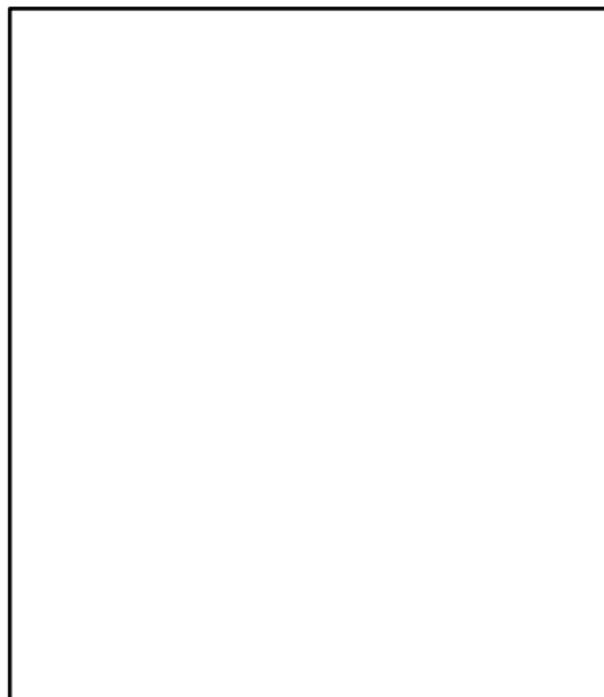
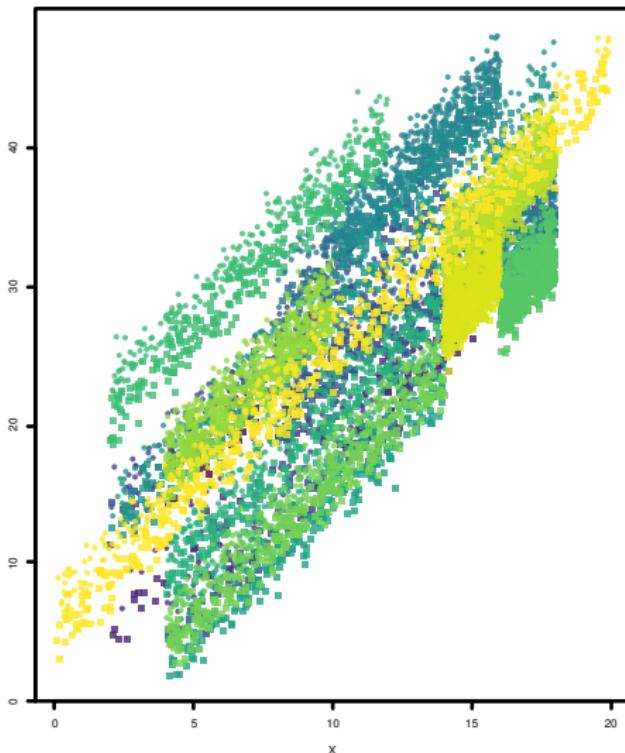


Removing site



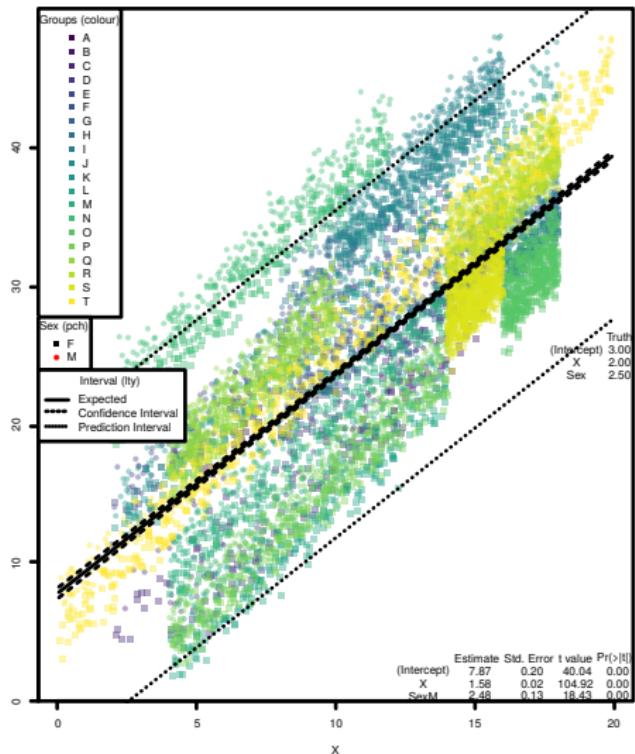
Summary

Accounting for multiple studies

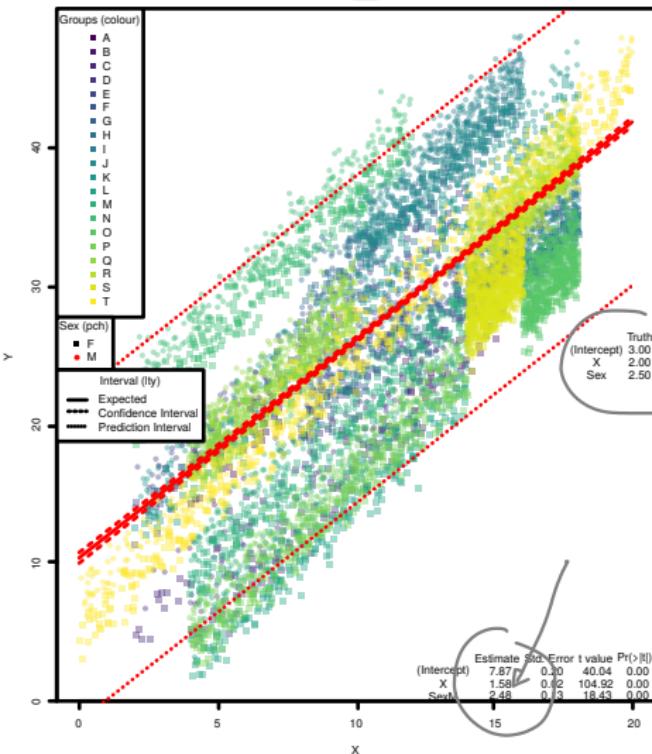


Accounting for multiple studies

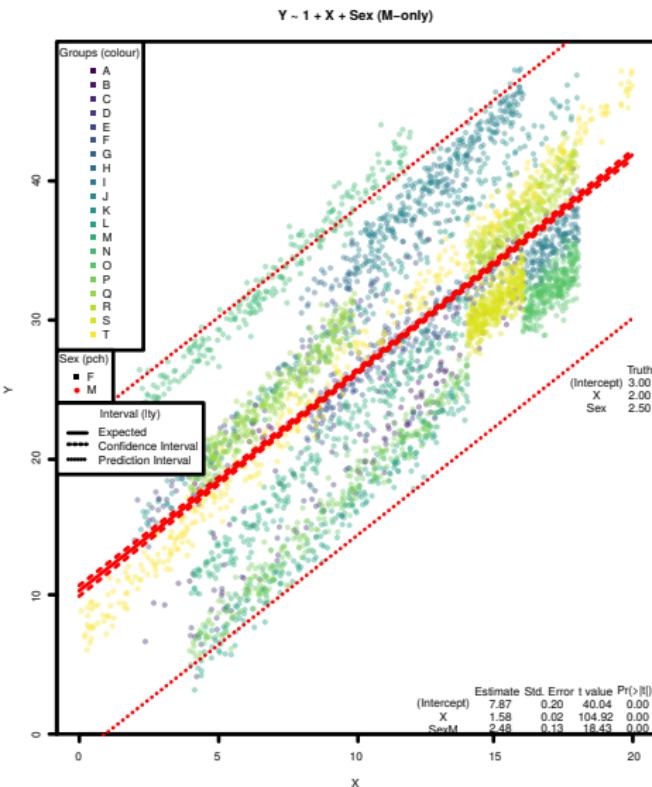
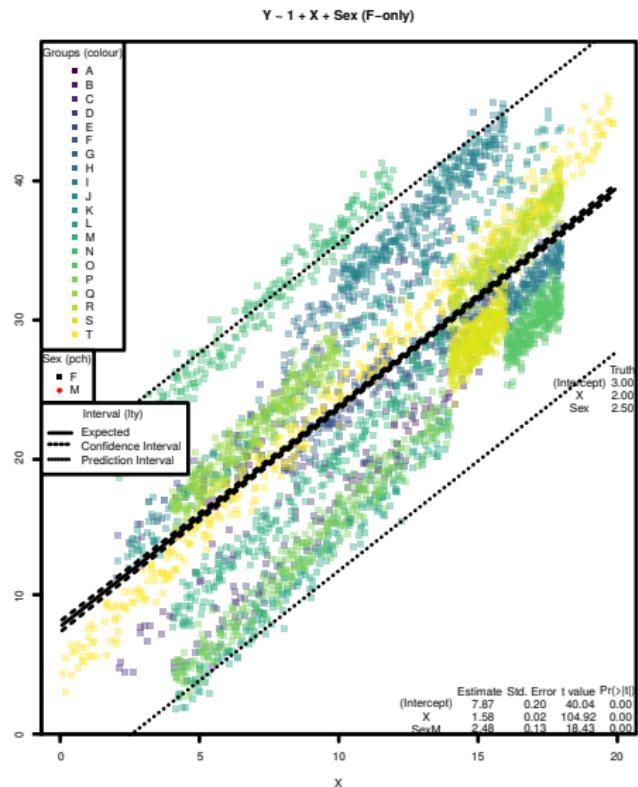
$Y \sim 1 + X + Sex$ (all data)



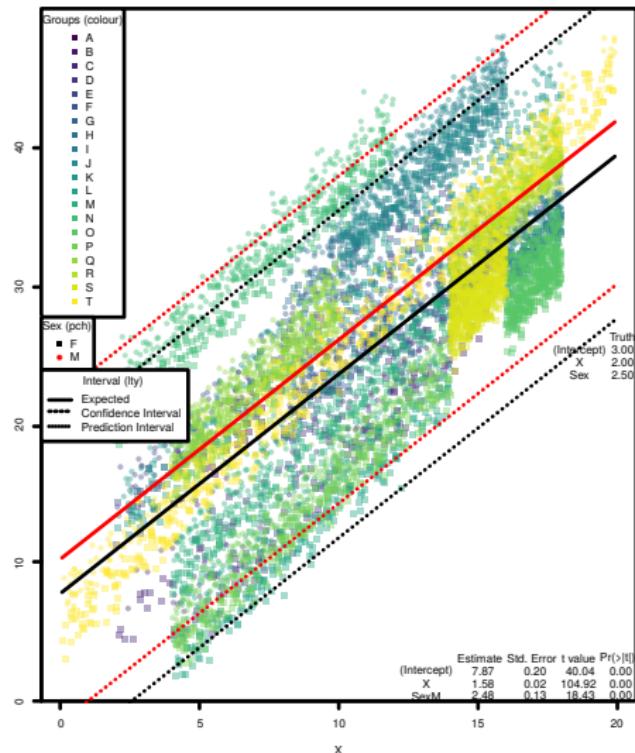
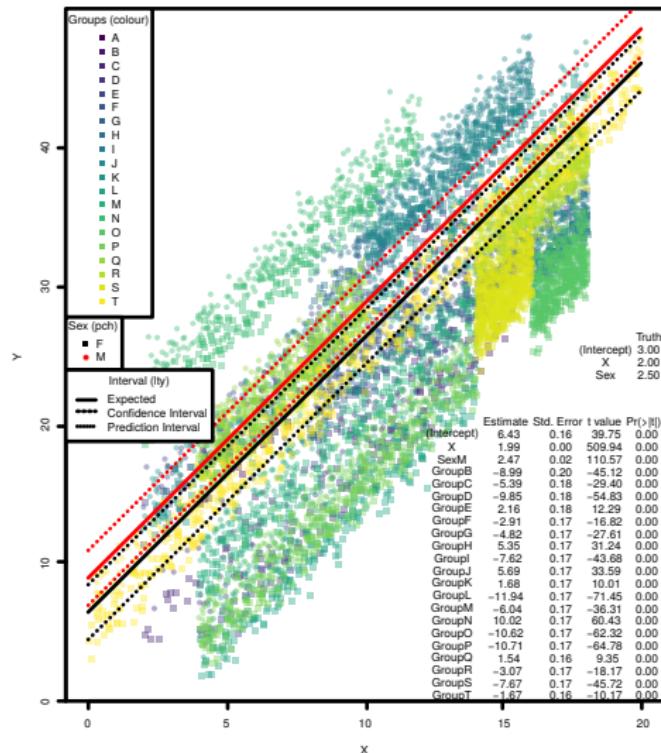
$Y \sim 1 + X + Sex$ (all data)



Accounting for multiple studies



Accounting for multiple studies

 $Y \sim 1 + X + \text{Sex}$  $Y \sim 1 + X + \text{Sex} + \text{Group (for Group=A)}$ 

Site?

Example 1

Identifiability

Example 2

Example 3

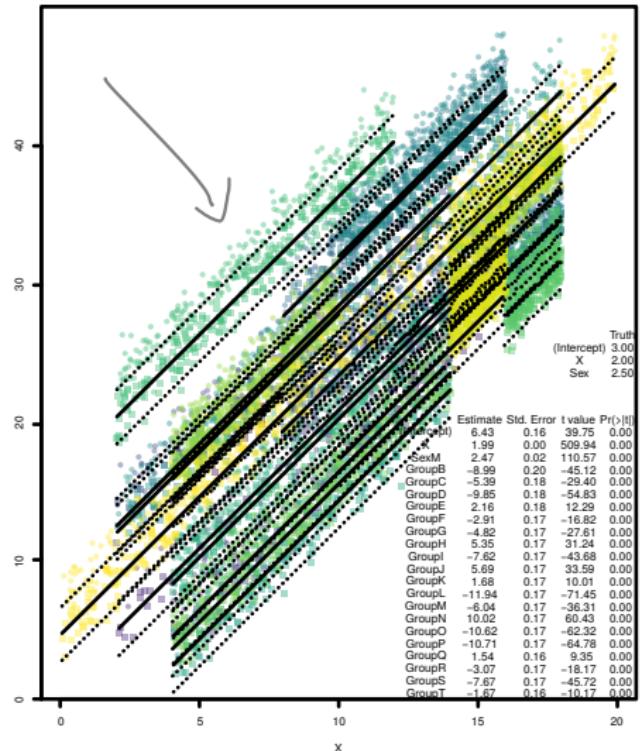
Example 4

Removing site

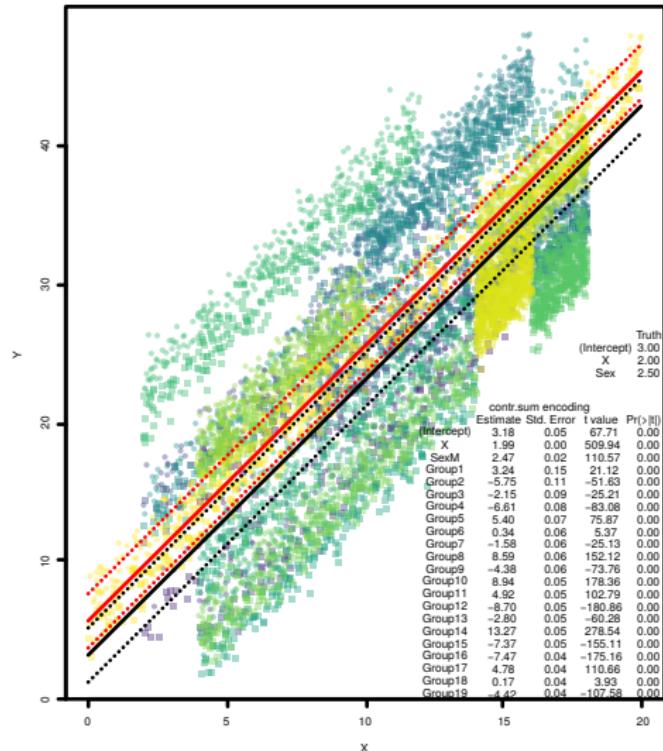
Summary

Accounting for multiple studies

$Y \sim 1 + X + Sex + Group$ (Group-specific fits)



$Y \sim 1 + X + Sex + Group$ (Group-average fit)



Site?



Example 1



Identifiability



Example 2



Example 3



Example 4

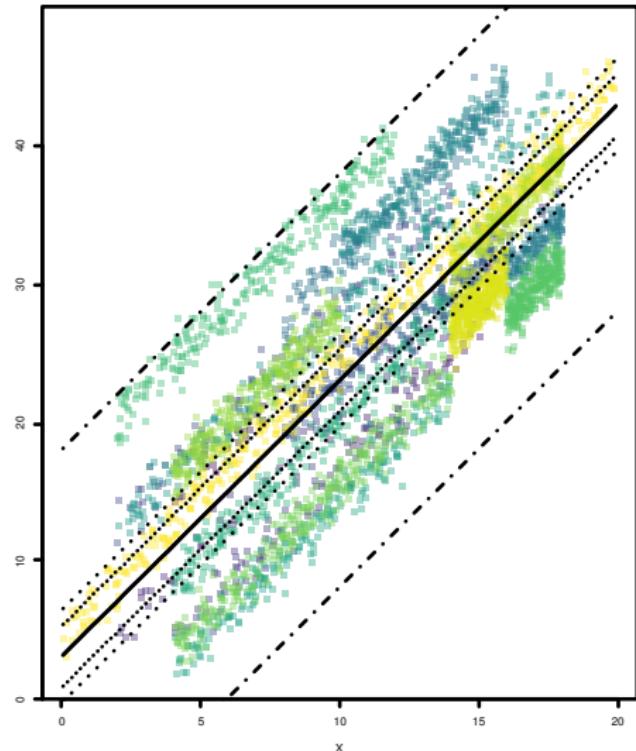
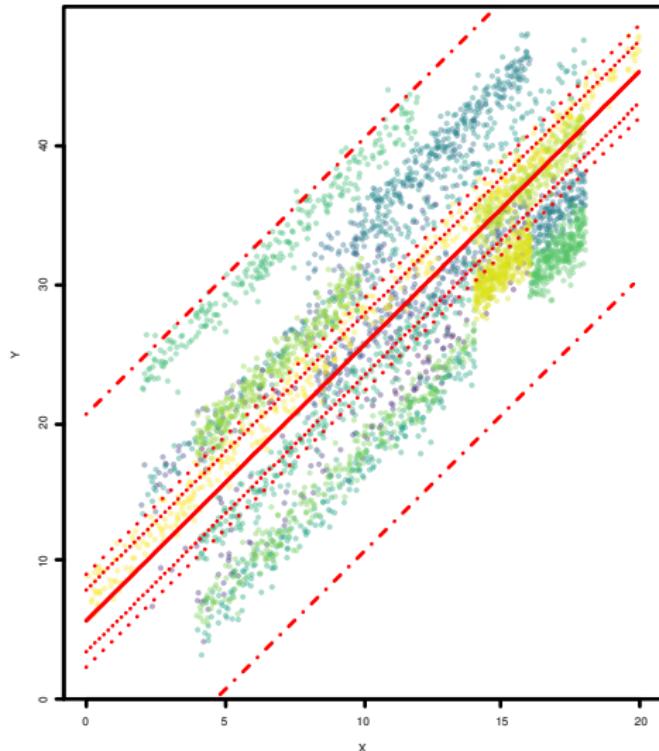


Removing site



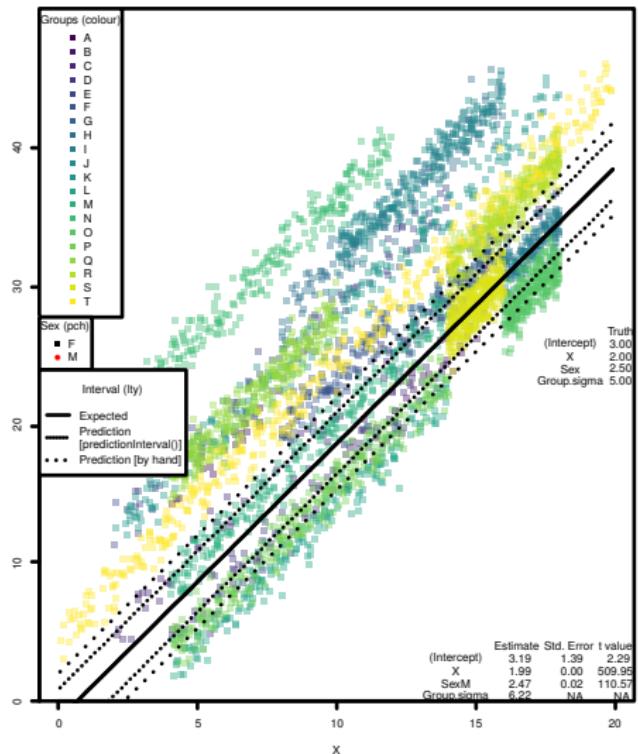
Summary

Accounting for multiple studies

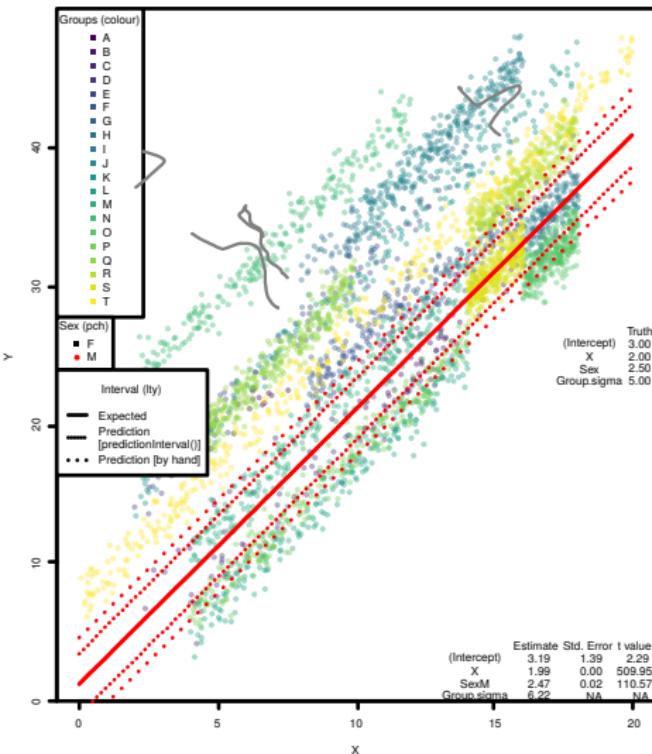
 $Y \sim 1 + X + \text{Sex} + (1|\text{Group})$ [for new/unknown Group] $Y \sim 1 + X + \text{Sex} + (1|\text{Group})$ [for new/unknown Group]

Accounting for multiple studies

$Y \sim 1 + X + Sex + (1|Group)$ [for Known Group=S]



$Y \sim 1 + X + Sex + (1|Group)$ [for Known Group=S]



All-in-one or multi-stage analysis

- Do you need “site-adjusted” measures as inputs into another analysis
- Can you combined adjustment within an all-in-one analysis

ComBAT

The procedure for the estimation of the site parameters γ_{iv} and δ_{iv} uses Empirical Bayes, and is described in [Johnson et al. 2007] and Fortin et al. [2017]. The final ComBat-harmonized cortical thickness measurements are defined as

$$y_{ij\nu}^{\text{ComBat}} = y_{ij\nu} + \hat{\alpha}_\nu - \mathbf{X}_{ij}\hat{\beta}_\nu - \gamma_{iv}^* + \hat{\alpha}_\nu + \mathbf{X}_{ij}\hat{\beta}_\nu$$

Annotations:

- 1**: Red arrow pointing to the $\hat{\alpha}_\nu$ term.
- site-spécif. add**: Red annotation above the γ_{iv}^* term.
- data**: Red circle around the $y_{ij\nu}$ term.
- diff Var**: Yellow circle around the $\hat{\alpha}_\nu$ term.
- Sex, age, bioint, CRP**: Green circle around the $\mathbf{X}_{ij}\hat{\beta}_\nu$ term.
- same type softm brn**: Green annotation to the right of the equation.

Site?

Example 1

Identifiability

Example 2

Example 3

Example 4

Removing site

Summary

Summary

- As always, the question is important
- There are many methods/approaches