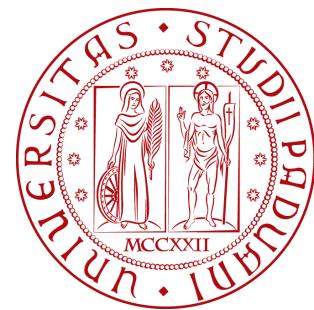


# Are we getting interactions wrong? The role of link functions in psychological research



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

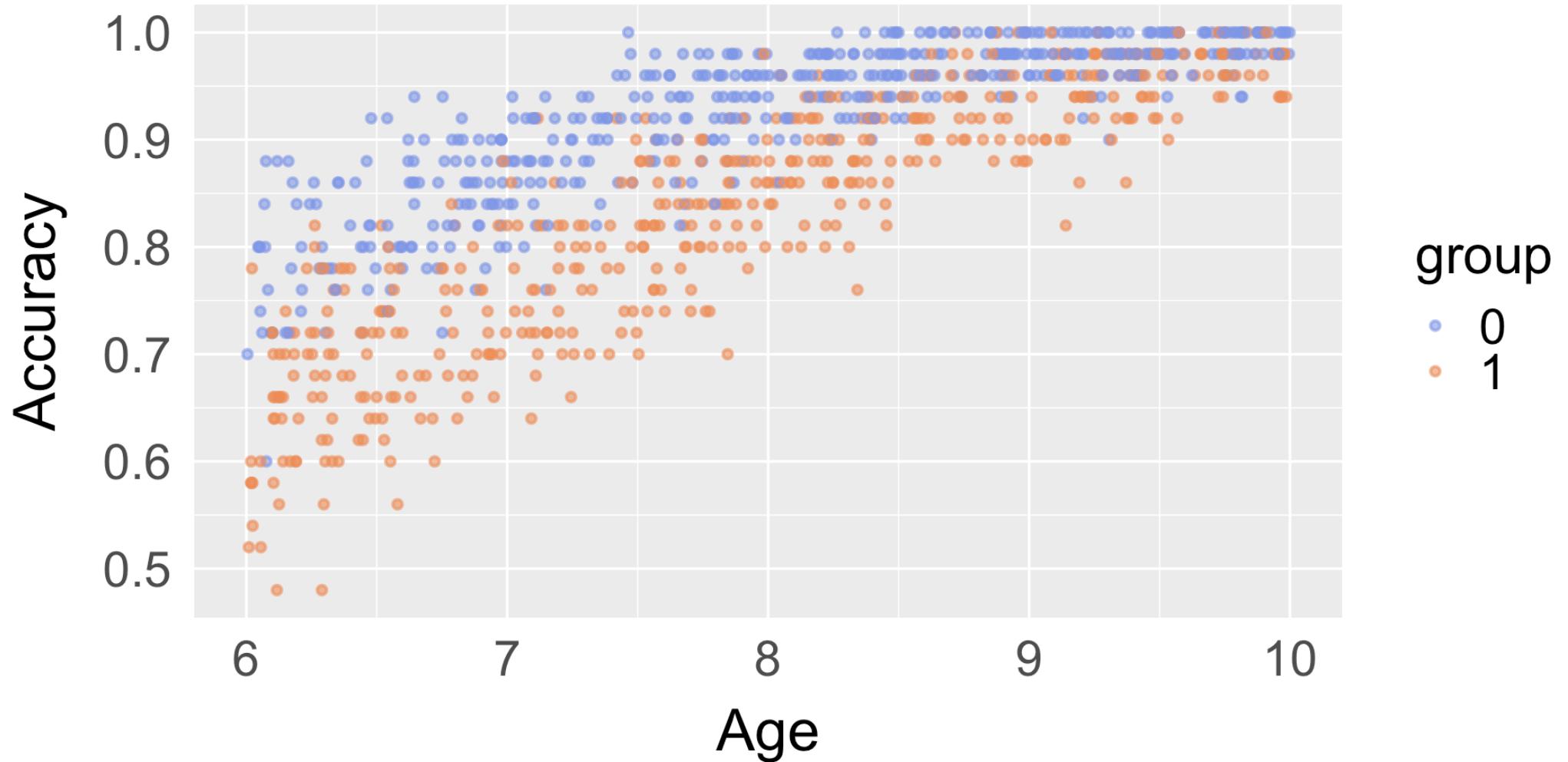
Laura Sità, Margherita Calderan, Tommaso Feraco,  
Filippo Gambarota, Enrico Toffalini

# Our dataset

- 1,000 subjects
  - 500 typically developing children (**group = 0**)
  - 500 children with dyslexia (**group = 1**)
- 50 trials per participant
- Independent variable 1: **age** (in years)
- Independent variable 2: **group**
- Dependent variable: **accuracy** in a TRUE/FALSE task



# Our dataset



# Building the model

Key choices:

- **family**

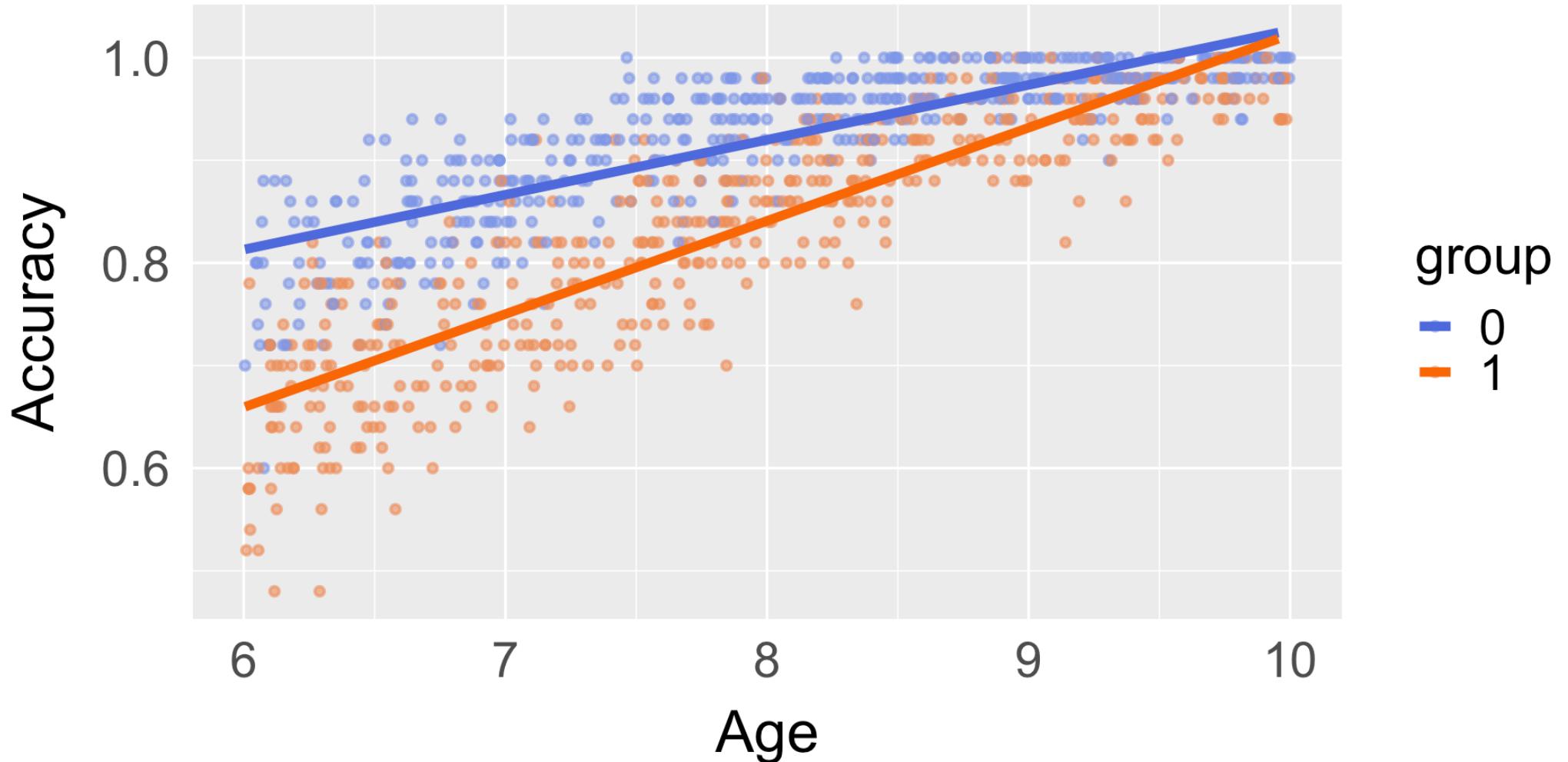
specifies the response distribution and its valid range  
(e.g., unbounded, [0, 1], counts)

- **link function**

maps the linear predictor  $\beta_0 + \beta_1 \cdot age + \beta_2 \cdot group$   
onto the scale of the response variable  $Y$

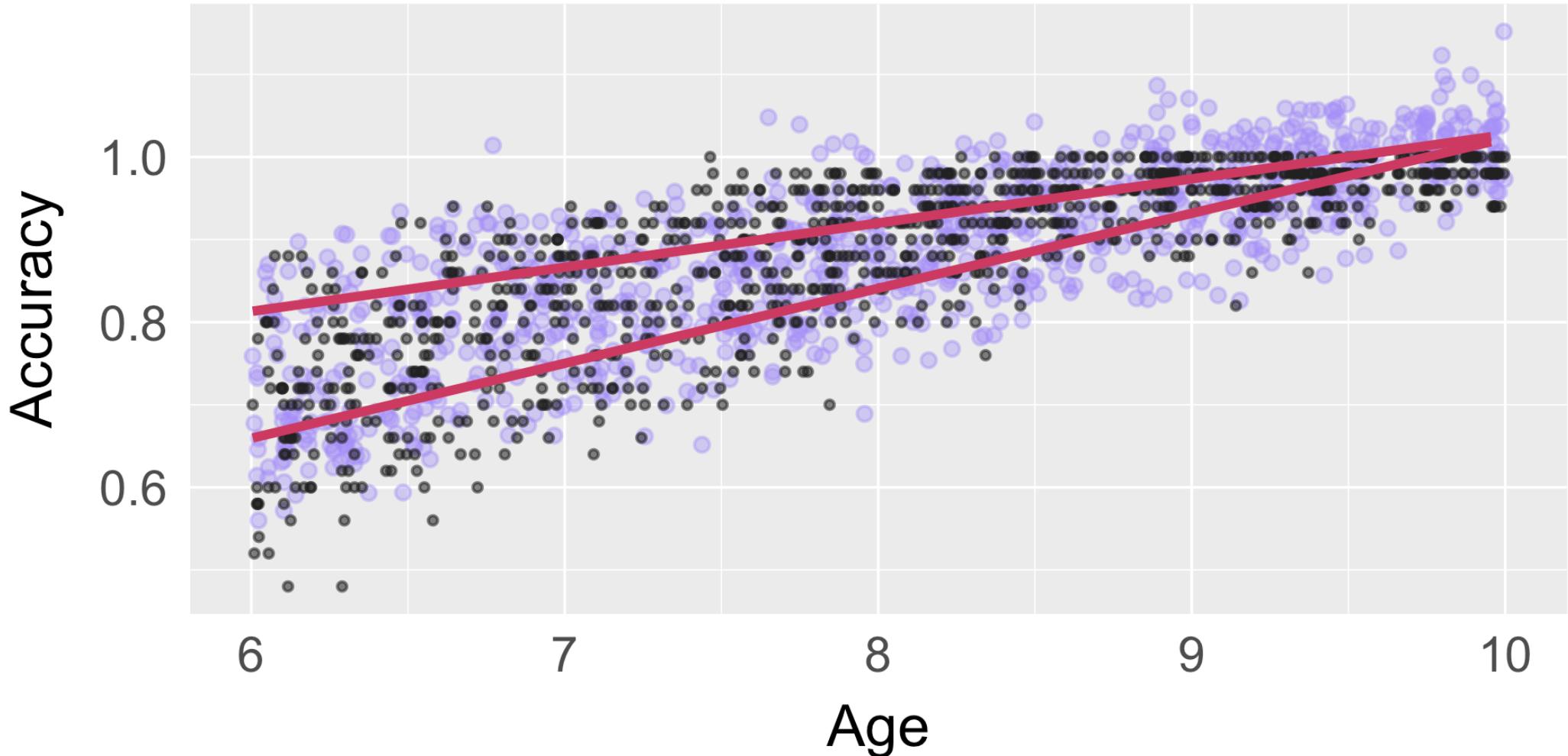
# Linear model

**family=gaussian(link="identity")**



# Predictive check

New predicted values fall outside the valid range for accuracy [0,1]



## **family=gaussian(link="identity")**

A **positive** interaction emerges

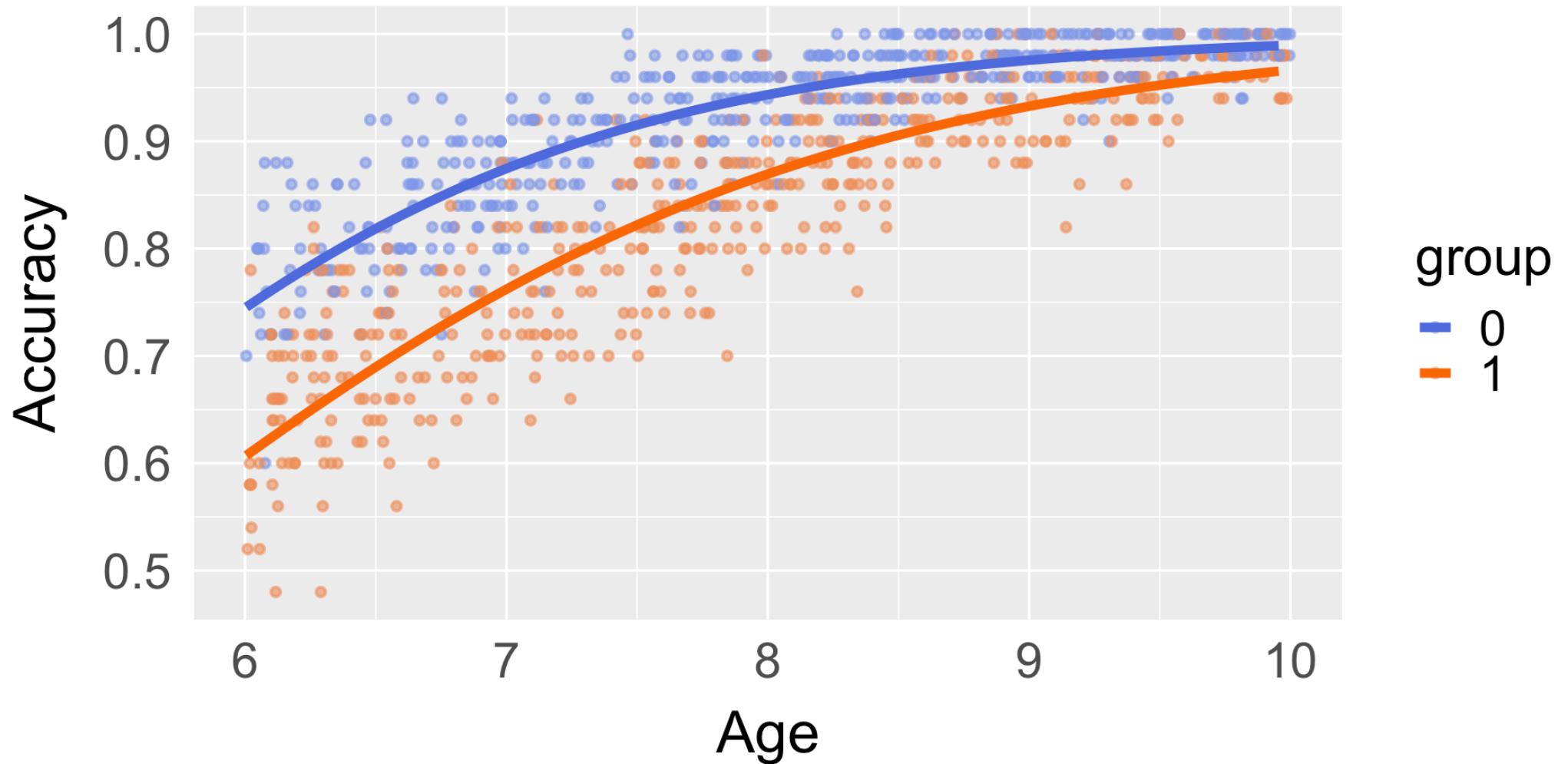
```
1 fit = glm(accuracy ~ age*group, family=gaussian(link="identity"), dat  
2 summary(fit)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	0.492140	0.015673	31.40	<2e-16	***
age	0.053481	0.001919	27.87	<2e-16	***
group1	-0.377555	0.021786	-17.33	<2e-16	***
age:group1	0.037318	0.002698	13.83	<2e-16	***

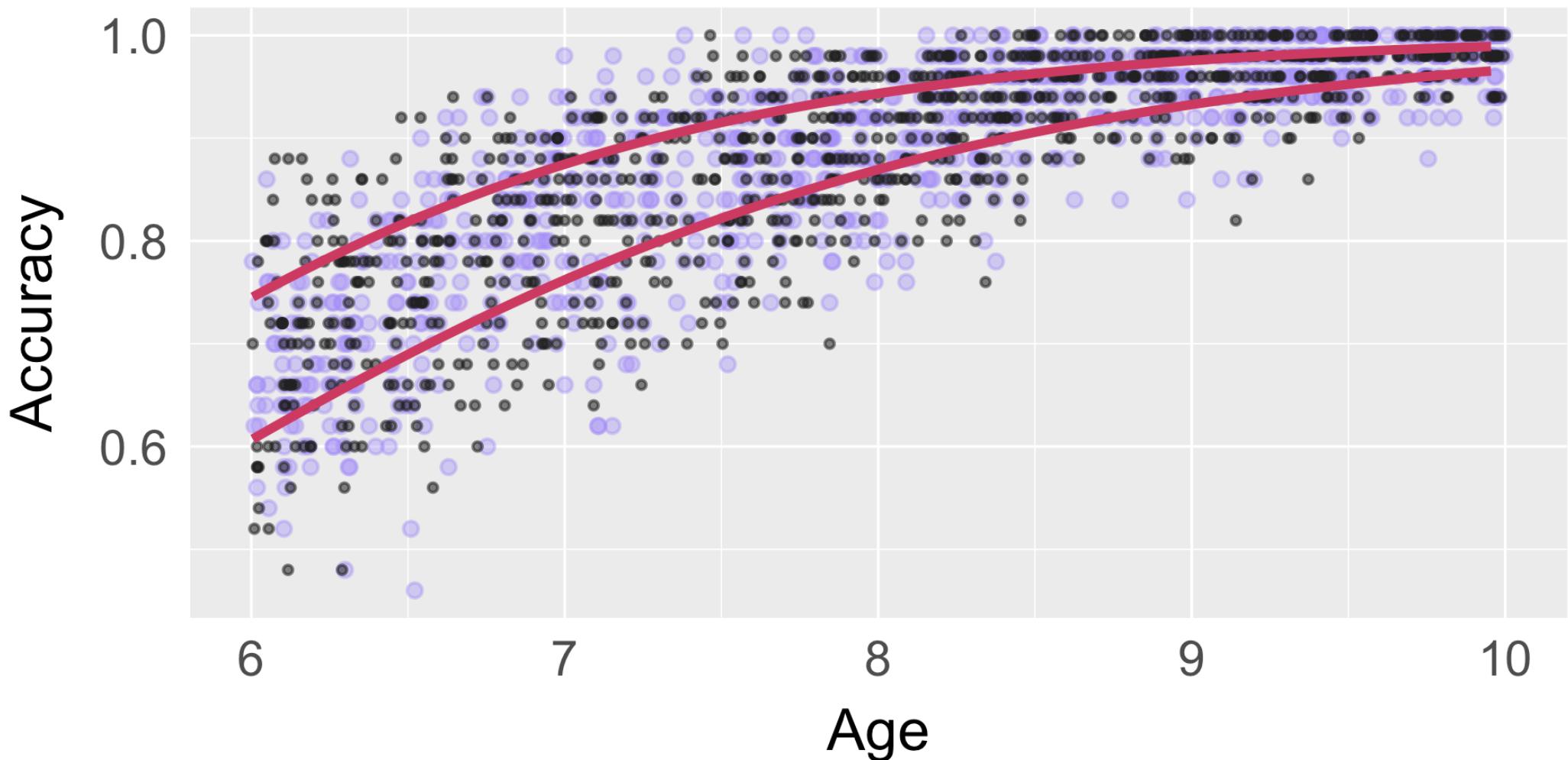
# Logistic regression model

`family=binomial(link="logit")`

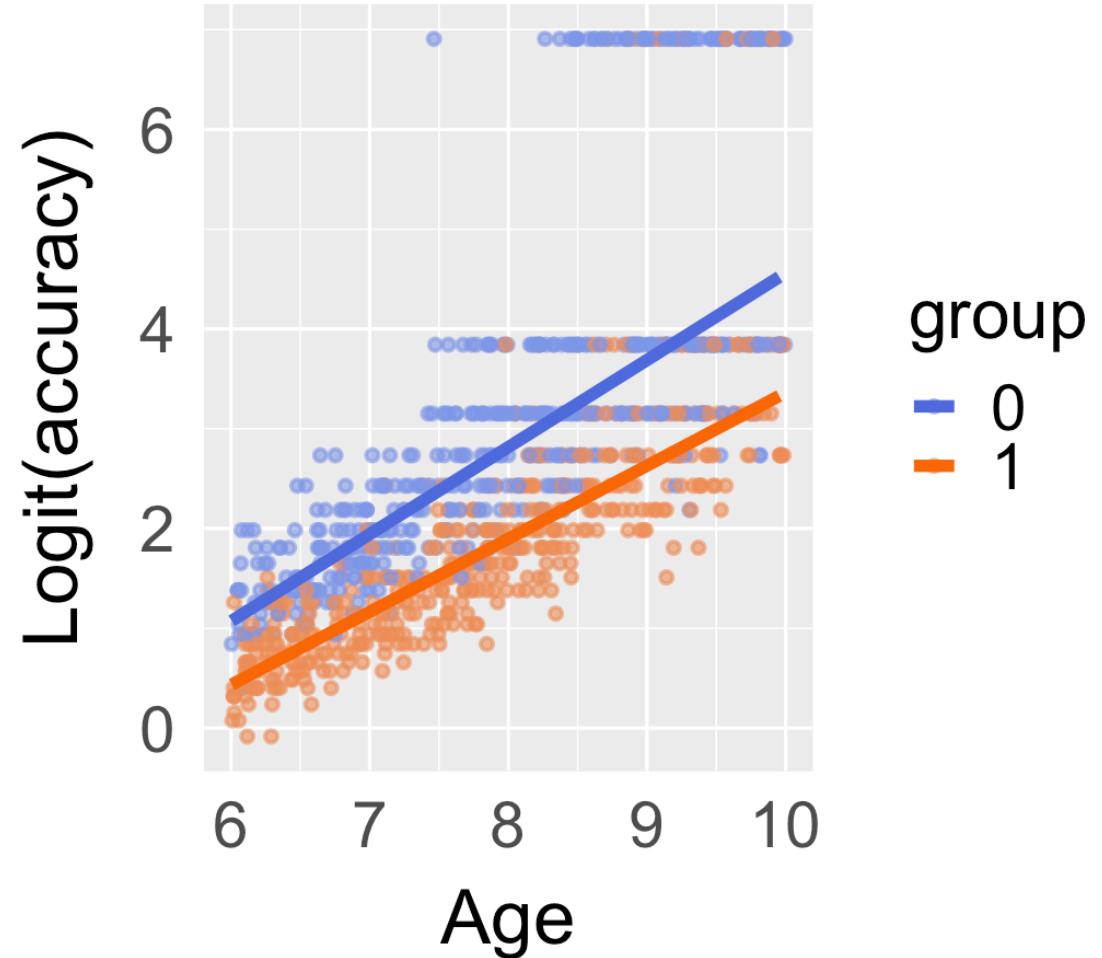
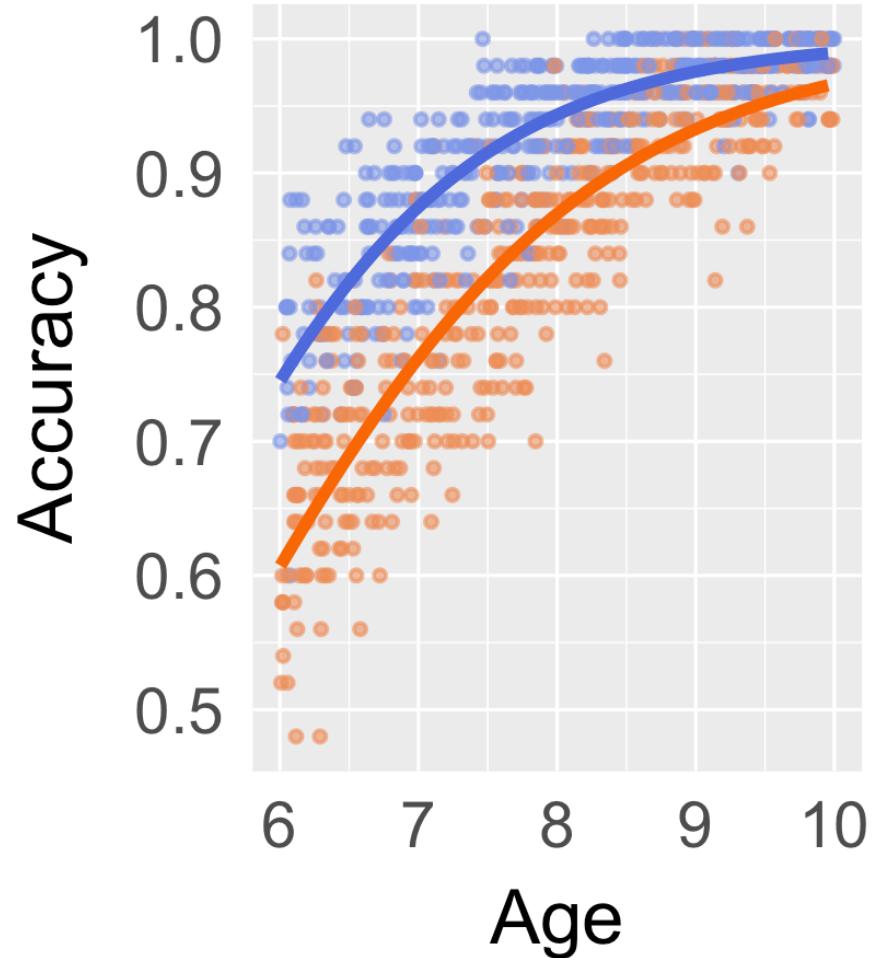


# Predictive check

New predicted values fall within the valid range for accuracy [0,1]



```
family=binomial(link="logit")
```



**family=binomial(link="logit")**

A **negative** interaction emerges

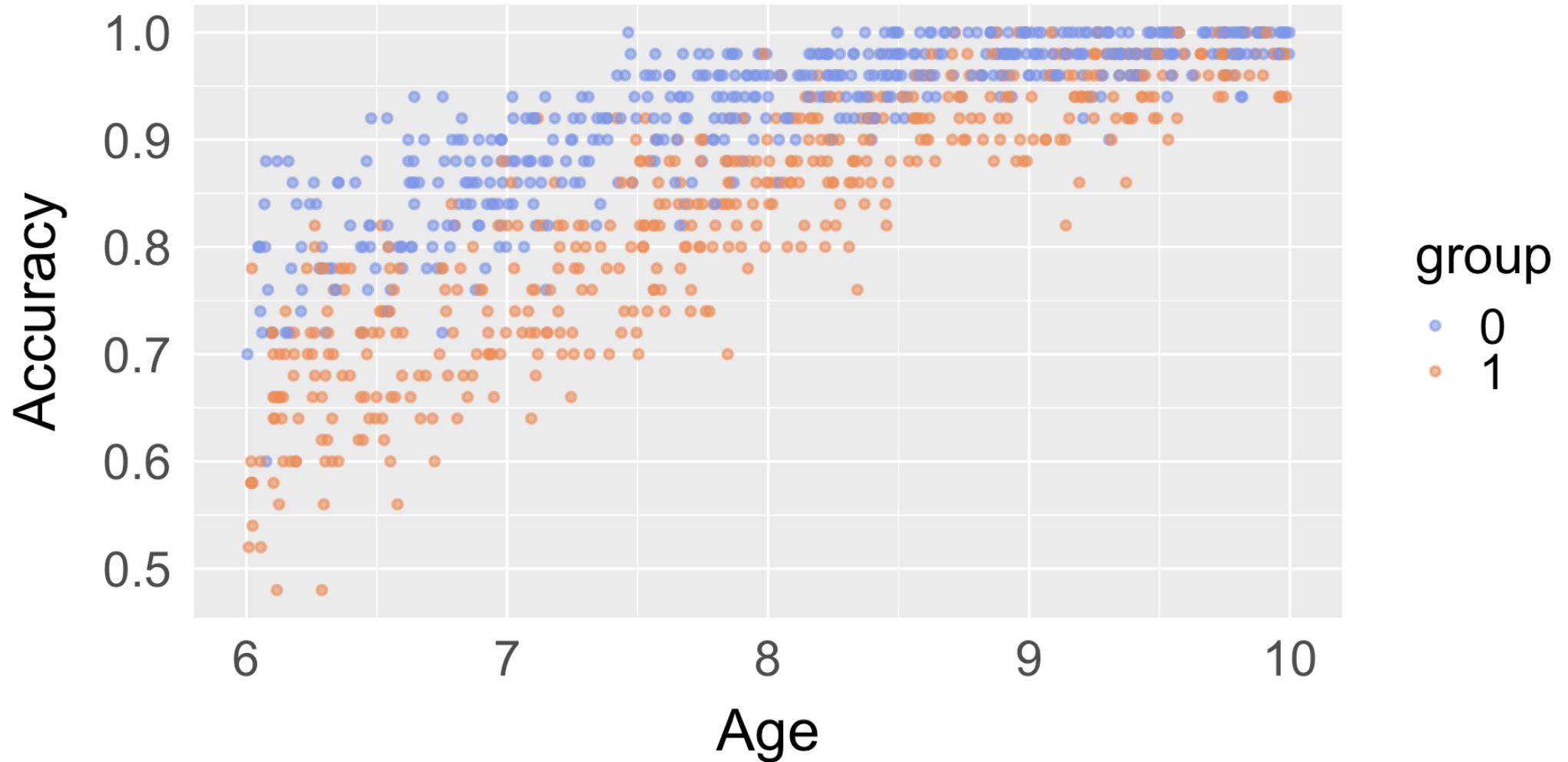
```
1 fit = glm(accuracy ~ age*group, data=d, family=binomial(link="logit"))
2 summary(fit)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-4.16402	0.19030	-21.881	< 2e-16	***
age	0.87240	0.02599	33.573	< 2e-16	***
group1	0.20253	0.23045	0.879	0.379	
age:group1	-0.14010	0.03139	-4.463	8.07e-06	***

# The appropriate model

# The dataset...



# ... was actually simulated

```
1 set.seed(123)
2
3 k = 50
4 N = 1000
5 group = rbinom(N,1,.5)
6 age = runif(N,6,10)
7 eta = -6+1*age-1*group # linear predictor
8 probs = mafc.logit(.m = 2)$linkinv(eta)
9 accuracy = rbinom(n = N, size = k, prob = probs) / k
```

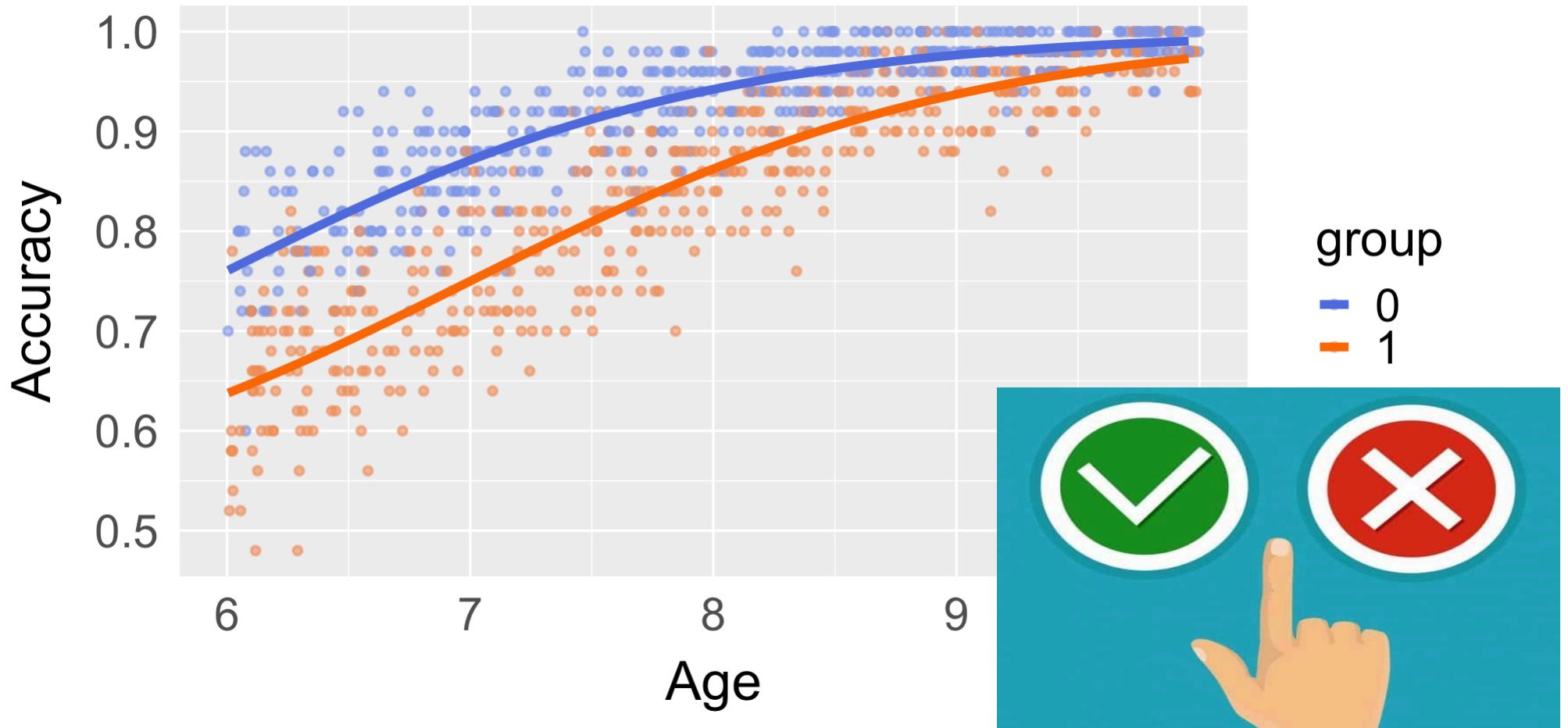
No interaction was simulated

Both models are detecting **an interaction that does not exist**

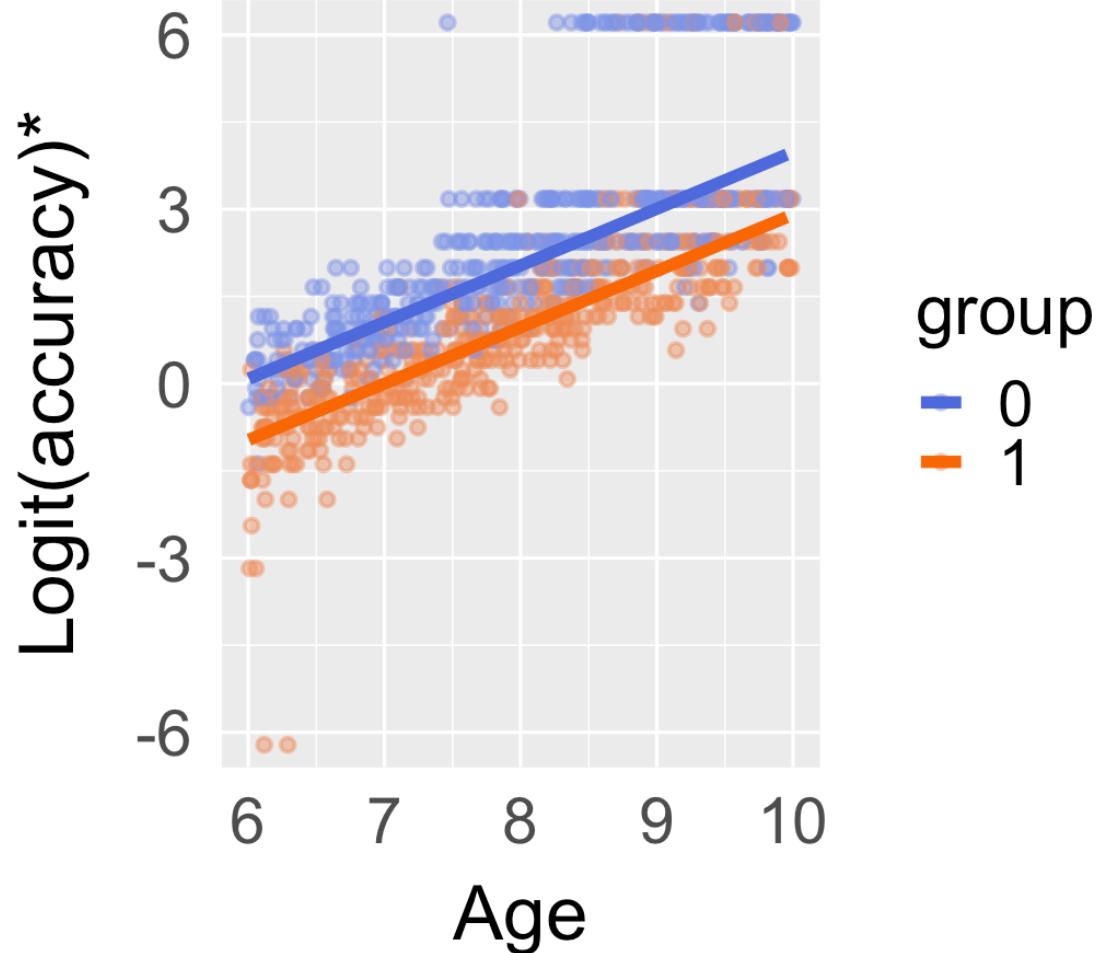
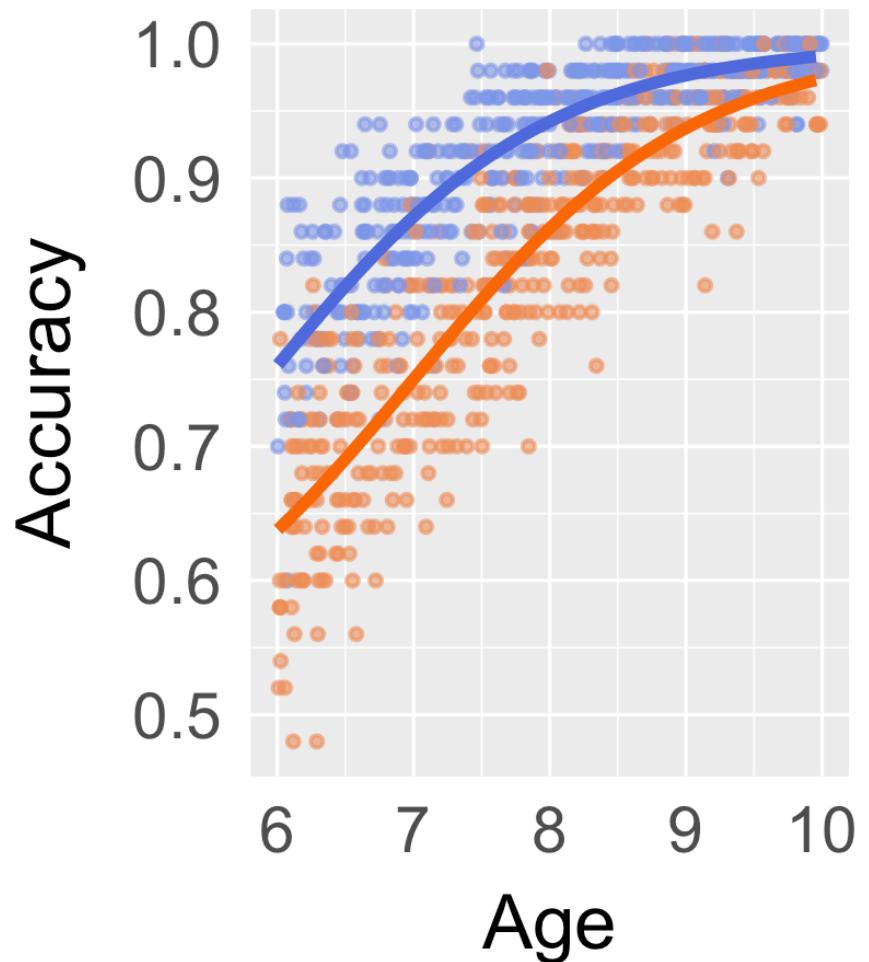
```
family=binomial(link=mafC.logit(.m=2))
```

To account for the 50% chance level in a TRUE/FALSE task:

**2 alternatives forced-choice logit link**



```
family=binomial(link=mafc.logit(.m=2))
```



\* Logit(accuracy) with 0.5 as lower bound

```
family=binomial(link=maf.c.logit(.m=2))
```

No interaction emerges, in line with how the data were generated

```
1 fit = glm(accuracy ~ age*group, data=d, family=binomial(link=maf.c.logit(.m=2)))
2 summary(fit)
```

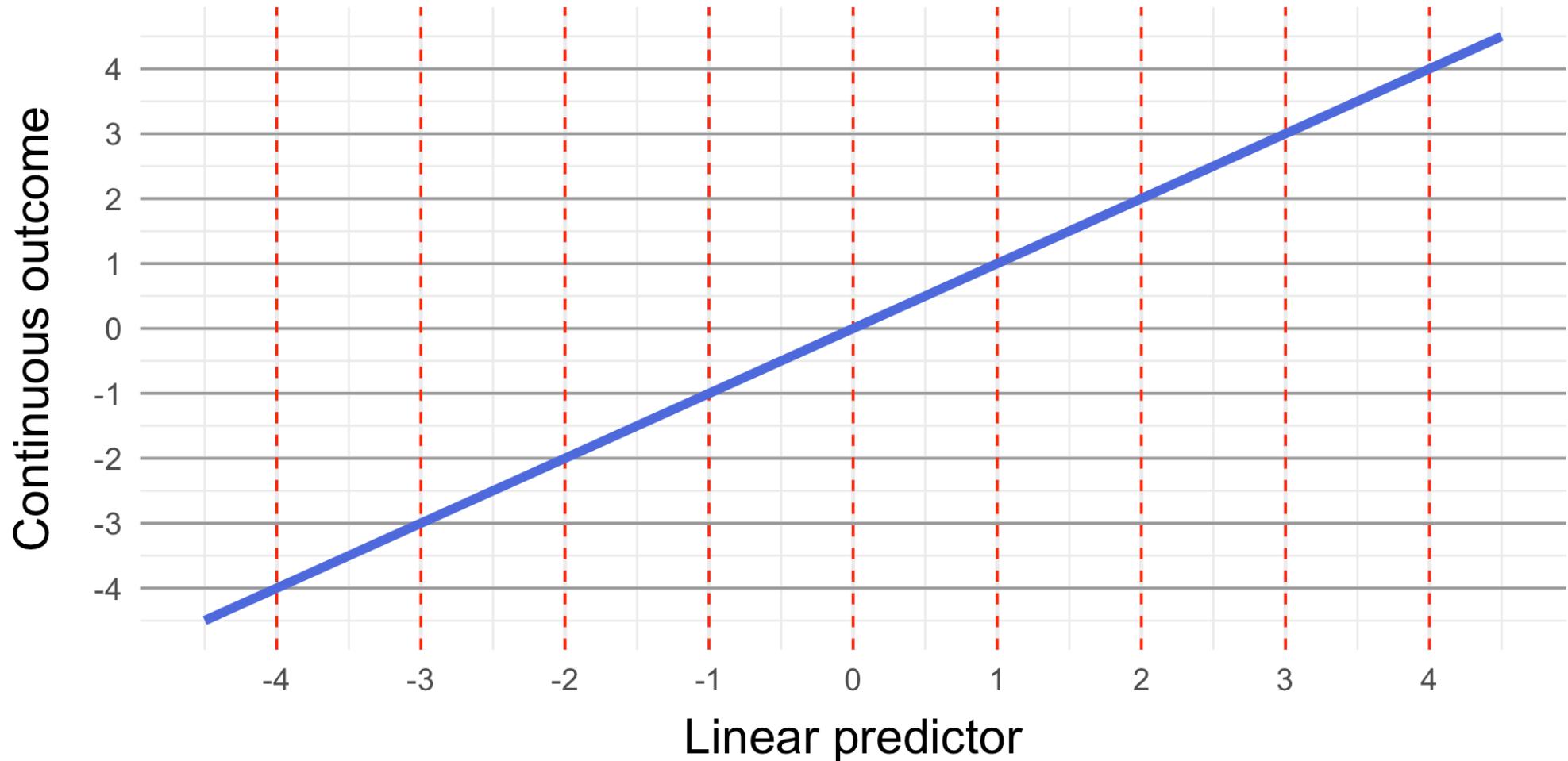
Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-5.773562	0.225527	-25.600	<2e-16	***
age	0.975904	0.030091	32.432	<2e-16	***
group1	-1.015102	0.299464	-3.390	0.0007	***
age:group1	-0.006119	0.039296	-0.156	0.8763	

# Why interactions

## **link="identity"**

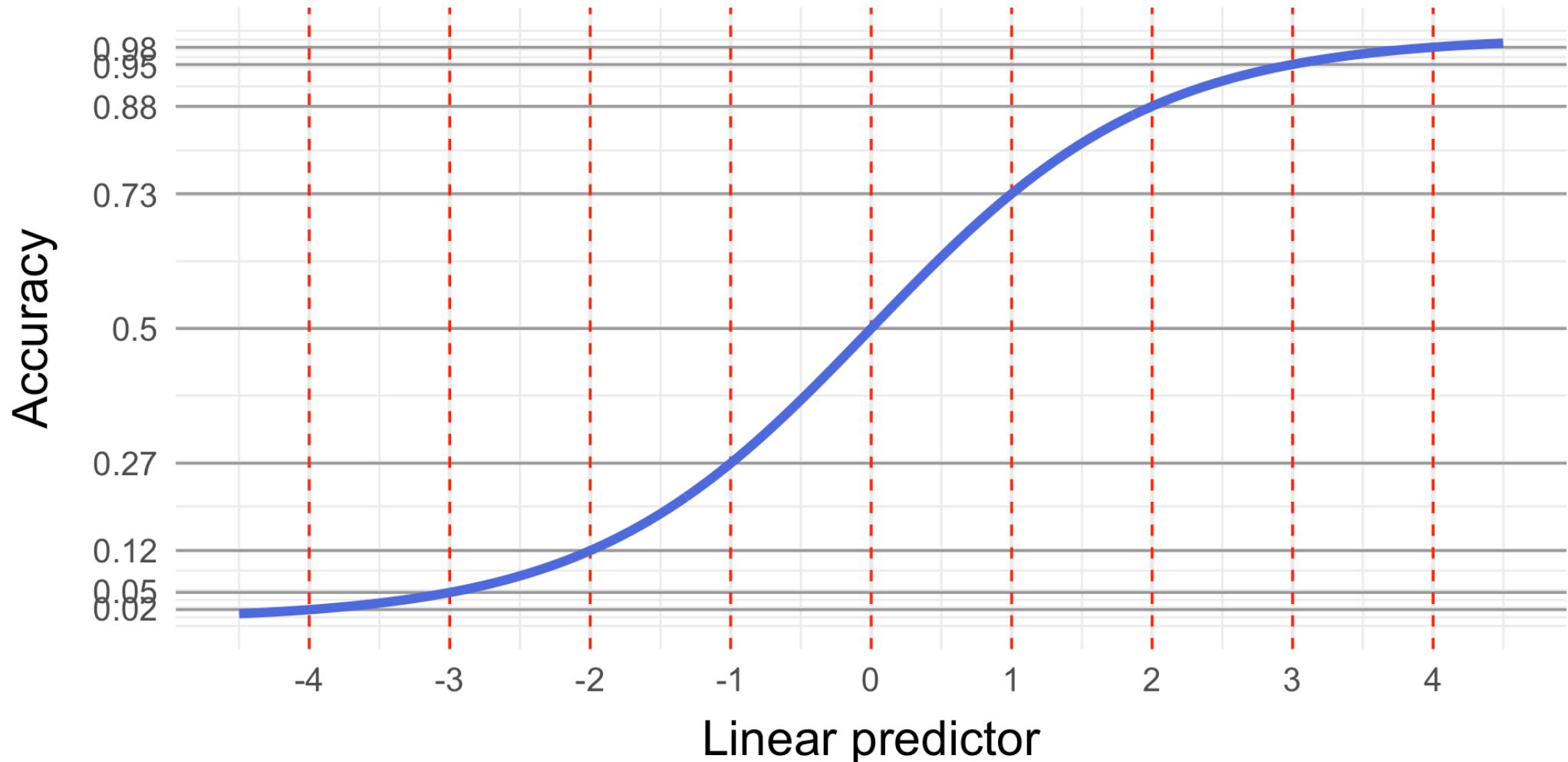
Equal intervals on X correspond to equal intervals on Y



In our example the linear predictor is  $\beta_0 + \beta_1 \cdot age + \beta_2 \cdot group$

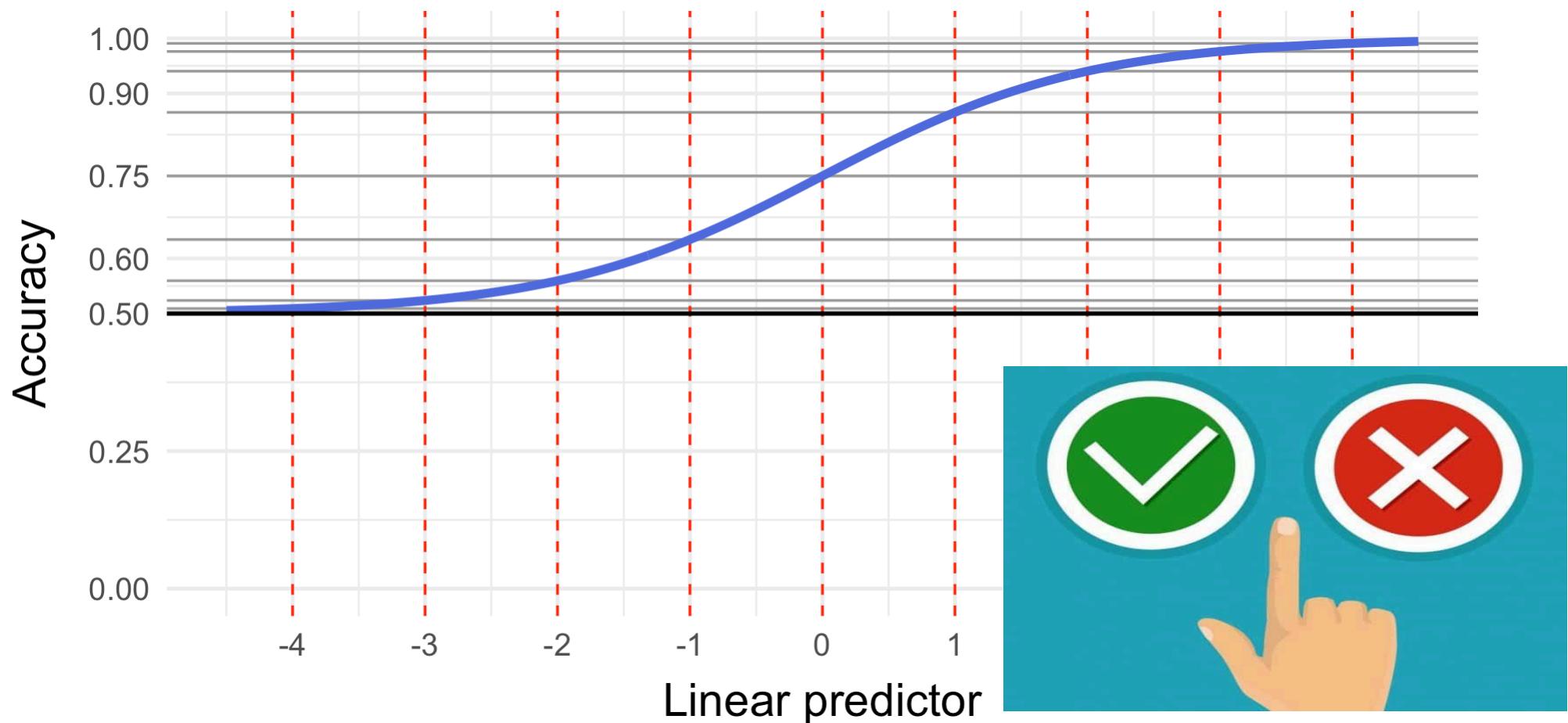
## [link="logit"](#)

Equal intervals on X correspond to equal intervals on Y **when on the logit scale**



## `link=mafcl.logit(2)`

Logit on [*chance level*, 1] instead of [0,1]



# Conclusions

Building a model means approximating the data-generating process  
(never observed directly in real data)

Key choices:



Family

Predicted values remain within the outcome's valid range



Link function

Different links change the outcome scale in different ways → estimated interactions change too

Wrong links can create spurious interactions

# Our systematic review of psychological research

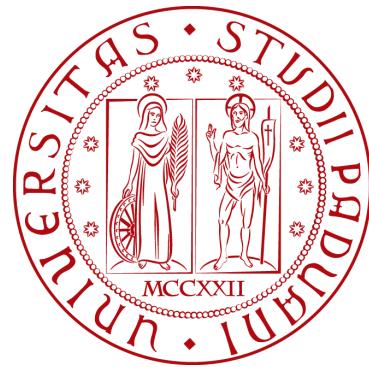
How often

- inappropriate link functions are used when testing interactions?
- do they lead to significant results?

# Materials & Contact

Data simulation, code and presentation are available on GitHub:  
[sitalaura / link-functions](https://github.com/sitalaura/link-functions)

Questions and feedbacks: [laura.sita@studenti.unipd.it](mailto:laura.sita@studenti.unipd.it)



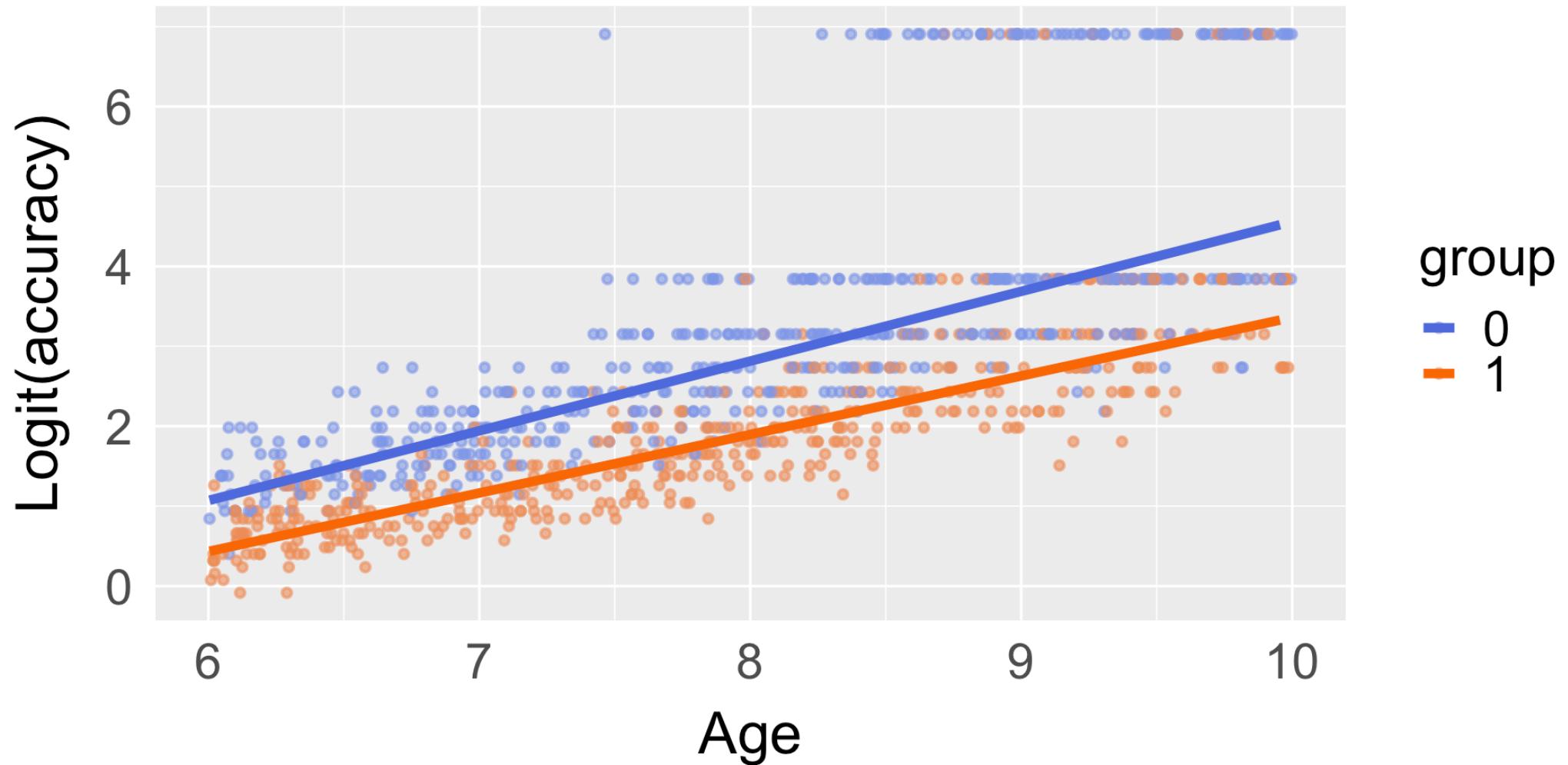
UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

# Bibliography

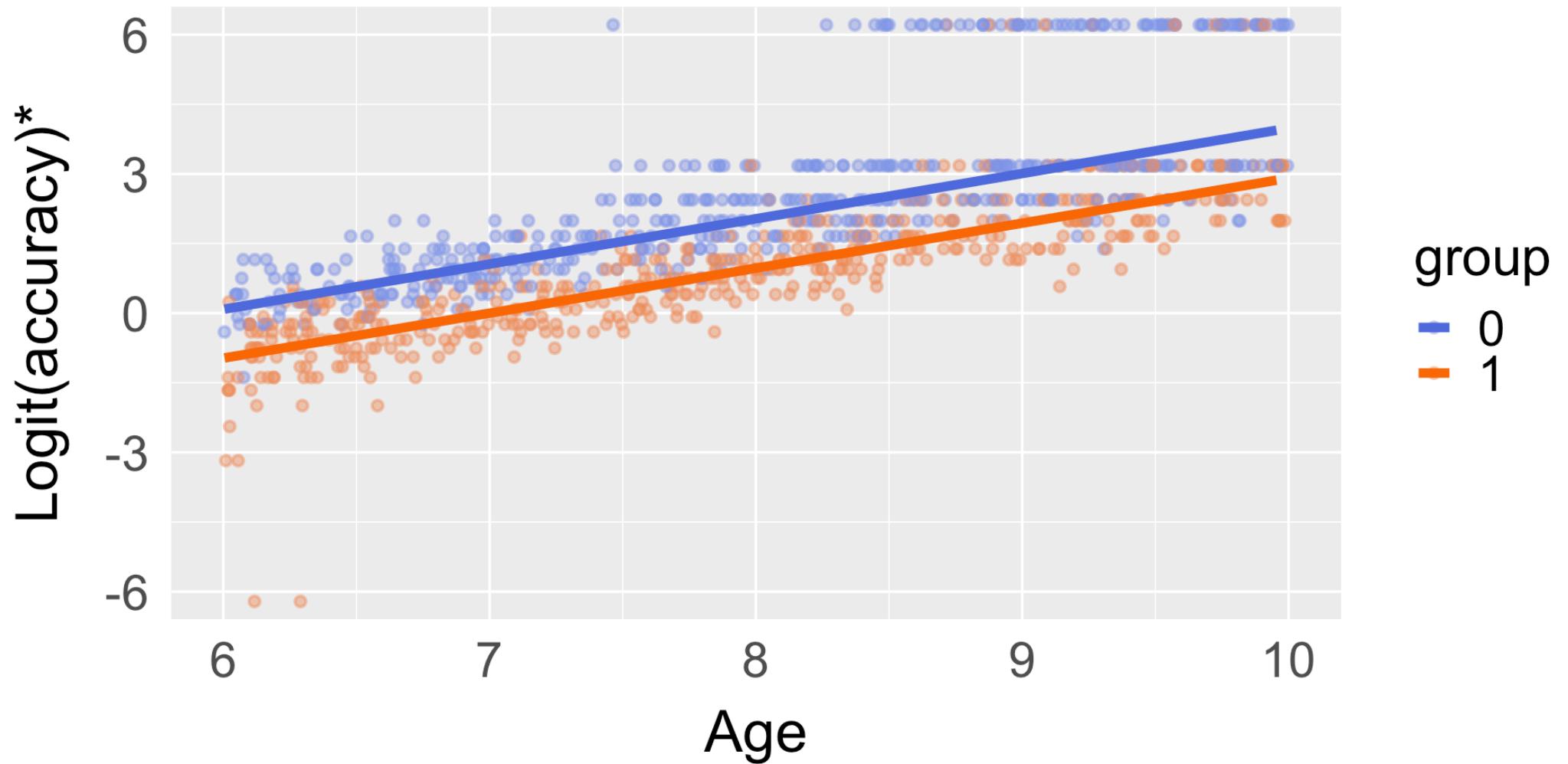
- Domingue, B. W., Kanopka, K., Trejo, S., Rhemtulla, M., & Tucker-Drob, E. M. (2024). Ubiquitous bias and false discovery due to model misspecification in analysis of statistical interactions: The role of the outcome's distribution and metric properties. *Psychological methods*, 29(6), 1164.
- Hardwicke, T. E., Thibault, R. T., Clarke, B., Moodie, N., Crüwell, S., Schiavone, S. R., Handcock, S. A., Nghiêm, K. A., Mody, F., Eerola, T., et al. (2024). Prevalence of transparent research practices in psychology: A cross-sectional study of empirical articles published in 2022. *Advances in Methods and Practices in Psychological Science*, 7 (4), 25152459241283477.
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong?. *Journal of Experimental Social Psychology*, 79, 328-348.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological bulletin*, 105(1), 156.

# Supplementary materials

# Logit(accuracy) with link=logit



# Logit(accuracy) with `link=mafC.logit(.m=2)`



\* Logit(accuracy) with 0.5 as lower bound