



Código  
Figuras  
Cosas que he olvidado decir



# Estimando curvas de adquisición léxica en la infancia

## Un modelo Bayesiano de TRI

- Gonzalo Garcia-Castro<sup>1</sup> @gongcastro
- Alicia Franco-Martínez<sup>2</sup> @AliciaFrancoXVE
- Cristina Rodríguez-Prada<sup>2</sup> @crodriguezprada
- Ignacio Castillejo<sup>2</sup> @IndaloNacho
- Núria Sebastian-Galles<sup>1</sup>



Universitat  
Pompeu Fabra  
Barcelona



Universidad Autónoma  
de Madrid

**Estimar la edad de adquisición de palabras en catalán y castellano en el Área Metropolitana de Barcelona**



**Edad de adquisición:** edad en la que una proporción umbral de la población de interés conoce o produce una palabra (e.j., 50%, 2 años)

Desarrollo comunicativo (10-32 meses)

Cumplimentado por una o dos personas a cargo del bebé

Inventario de palabras: ~400 ítems

- No / Comprende (10-16 meses)
- No / Produce (17-32 meses)

Adaptado a múltiples lenguas ([Wordbank](#))

Tamaño de vocabulario: suma del número de respuestas *Comprende* o *Produce*

1,600 ítems (798 catalán, 802 castellano)

Participantes asignados aleatoriamente a una de 4 versiones

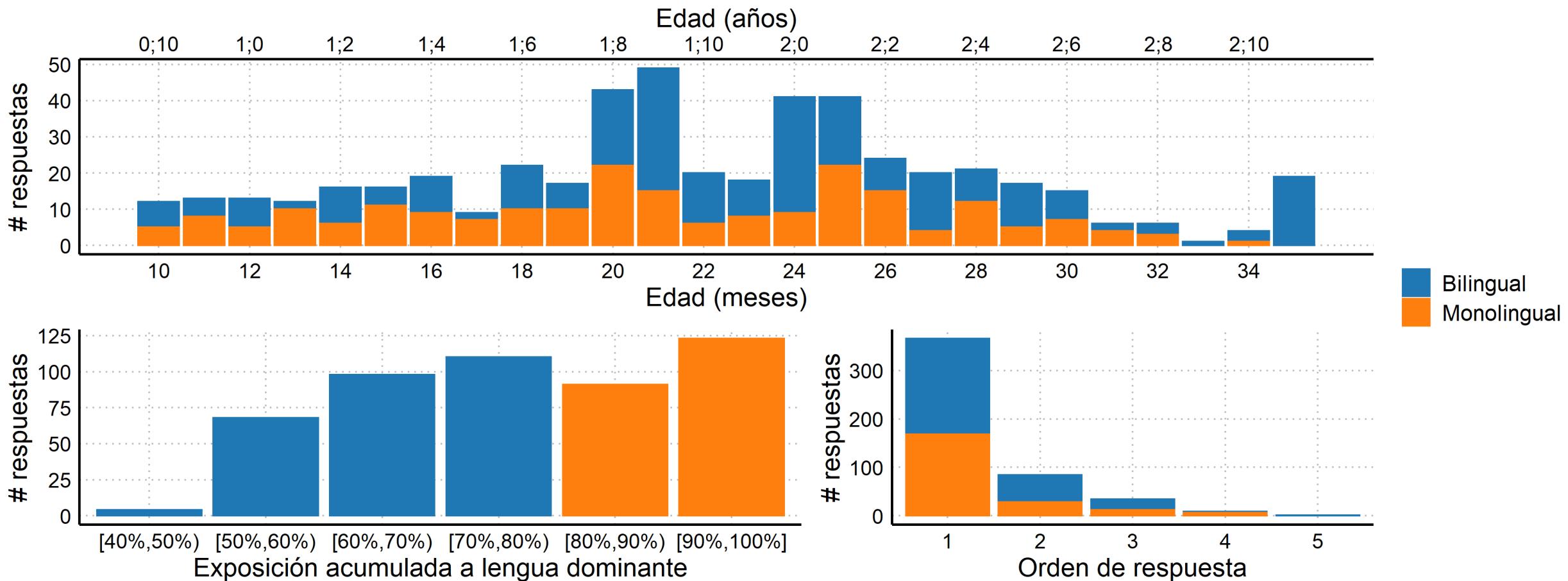
- ~560 items (280 cat, 280 spa)

Paquete de R: [{multilex}](#) [[GitHub](#)]



# Base de datos

417 participantes, 494 respuestas, 254,813 observaciones



## Variable de respuesta

**response**: *No/Comprende/Comprende y dice*

## Variables del participante:

**id**: identificador de participante (*grupos*)

**age**: edad en meses

**doe**: exposición a la lengua (proporción)

## Variables del ítem:

**te**: identificador de traducción  
(*grupos*)

**item**: identificador de ítem

**freq**: frecuencia léxica

**n\_phon**: número de fonemas

```
## # A tibble: 10 x 8
##   id      age  doe    te item      freq n_phon response
##   <chr>    <dbl> <dbl> <int> <chr>     <dbl> <int> <ord>
## 1 bilexicon_642 10.2  0.5     1 cat_pessigolles  4.05    9 Understands
## 2 bilexicon_642 10.2  0.5     1 spa_cosquillas  3.83    8 Understands
## 3 bilexicon_634 10.7  1       1 cat_pessigolles  4.05    9 No
## 4 bilexicon_634 10.7  0       1 spa_cosquillas  3.83    8 No
## 5 bilexicon_878 10.8  0       1 cat_pessigolles  4.05    9 No
## 6 bilexicon_878 10.8  1       1 spa_cosquillas  3.83    8 No
## 7 bilexicon_1604 10.8  0.4    1 cat_pessigolles  4.05    9 No
## 8 bilexicon_1604 10.8  0.6    1 spa_cosquillas  3.83    8 No
## 9 bilexicon_1602 11.2  0.05   1 cat_pessigolles  4.05    9 No
## 10 bilexicon_1602 11.2  0.95  1 spa_cosquillas  3.83    8 No
```

# Modelo

Ordinal: No → Comprende → Comprende y dice) (función de enlace *logit*)

Multinivel: observaciones agrupadas por participant (**id**) e ítem (**te**)

Bayesiano:  $\text{P}(\theta | y) \propto P(y|\theta) \times P(\theta)$

- $y$  = probabilidad de respuesta a categoría (No, Comprende, Comprende y dice)
- $\theta$  = el modelo y sus parámetros

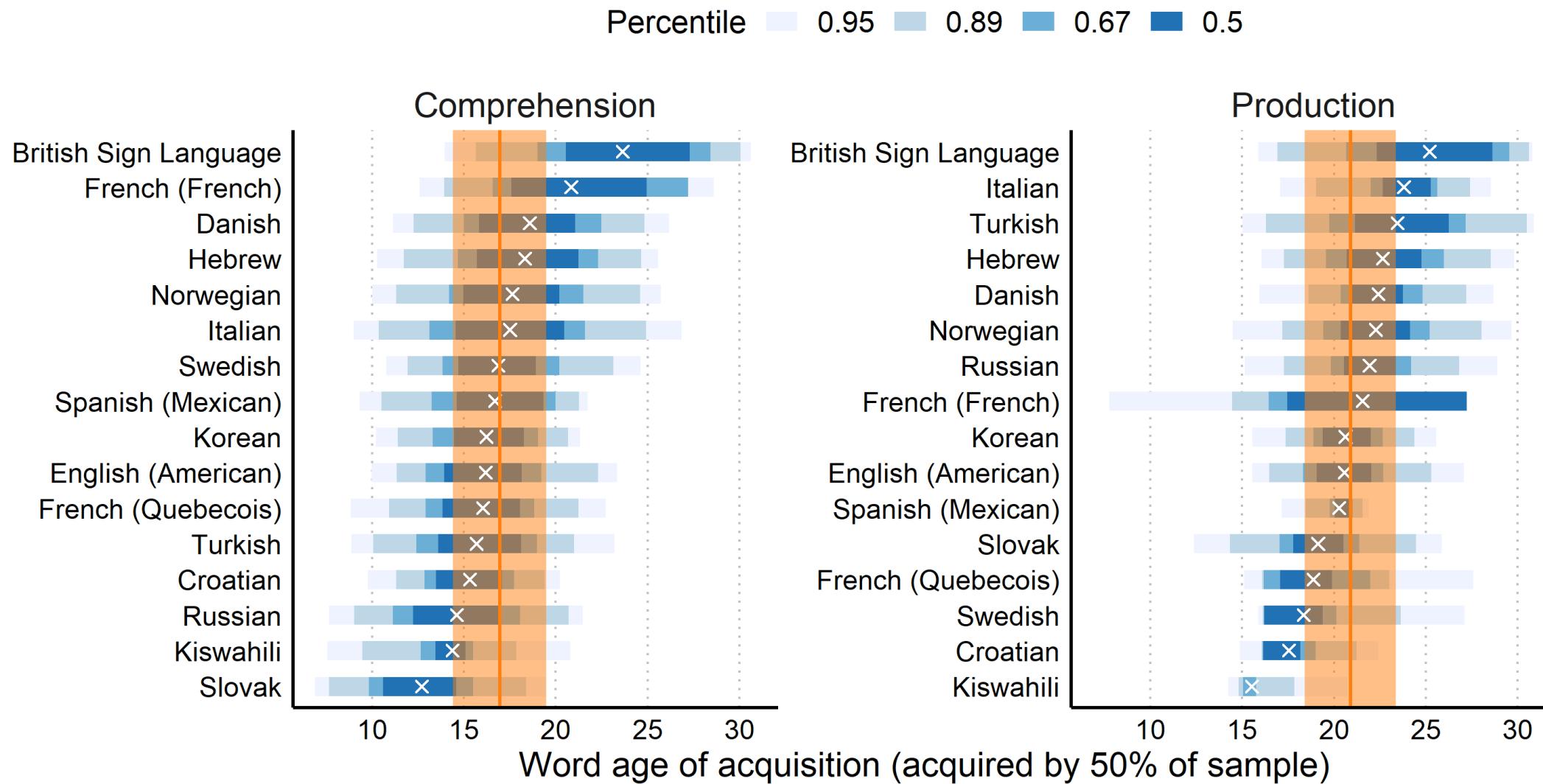
Implementación en {brms}/Stan:

```
response ~ age_std + freq_std + n_phon_std + doe_std +
(1 + age_std + freq_std + n_phon_std + doe_std | id) +
(1 + age_std + freq_std + n_phon_std + doe_std | te),
family = cratio(link = "logit") # cumulative, continuation ratio
)
```

# Distribución previa

# Distribution of ages of acquisition across languages

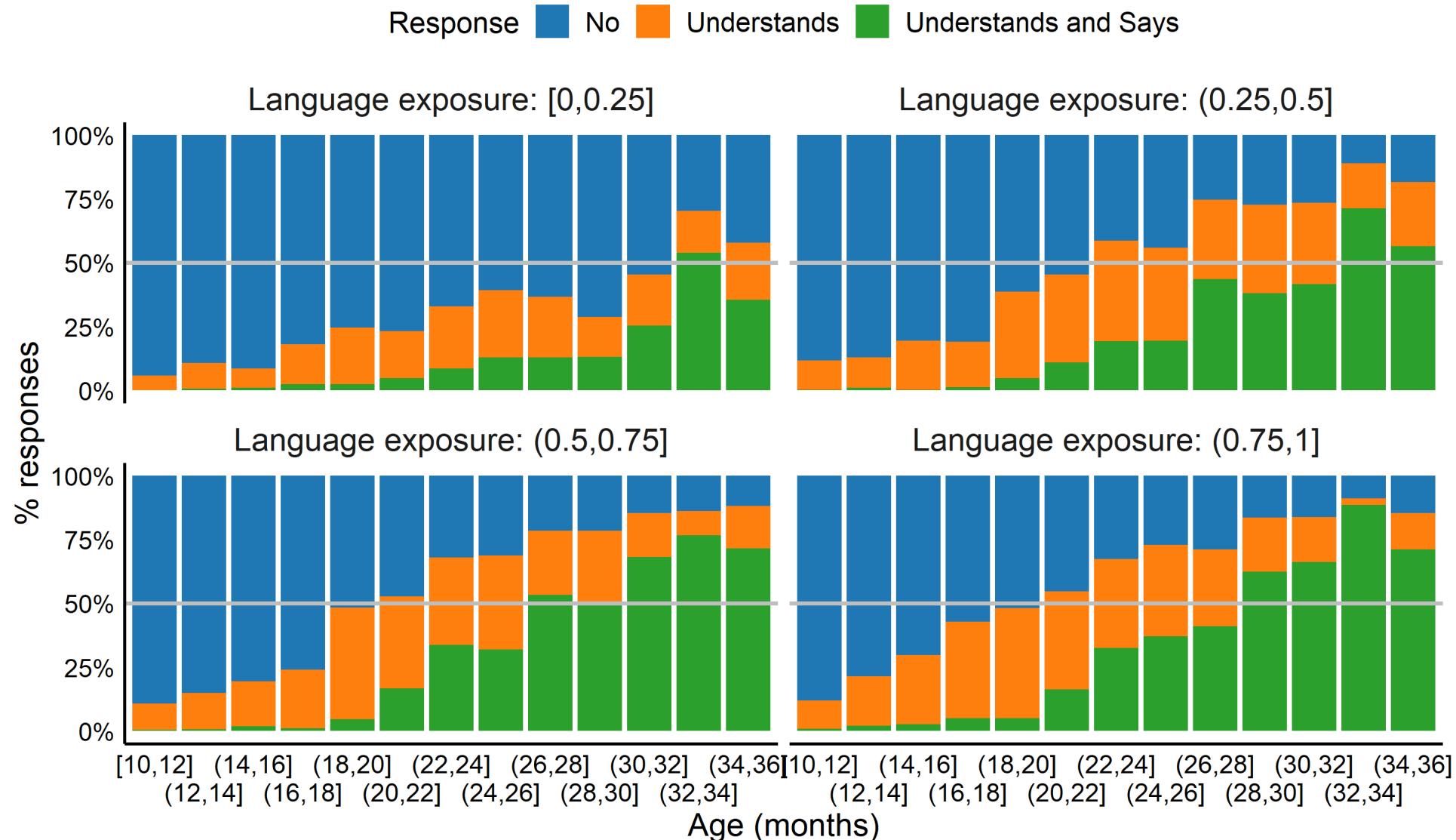
Orange lines and intervals indicate grand mean and SD



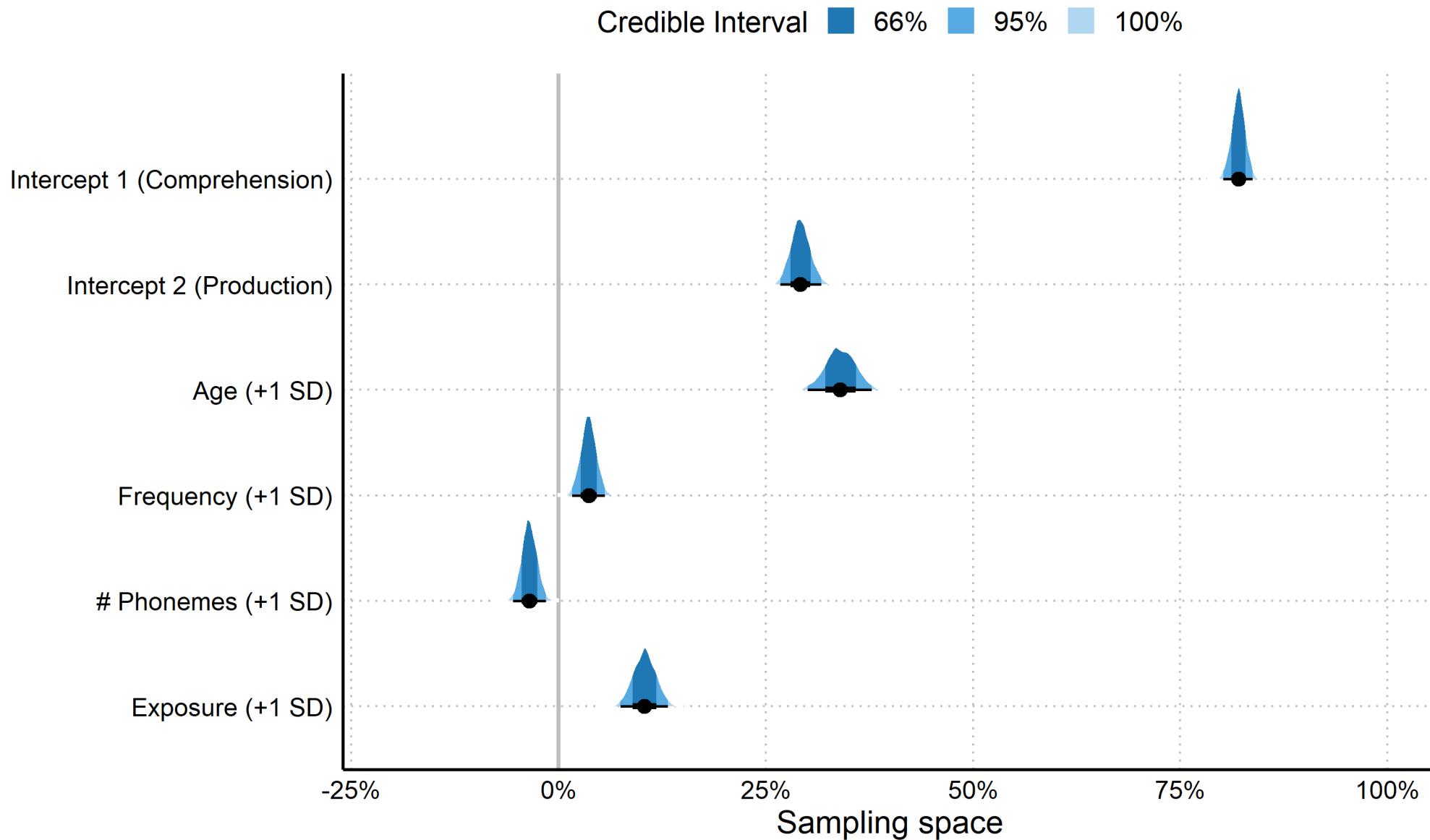
Source: Wordbank (see [https://mikabrv.io/aoa-prediction/aoa\\_estimation.html](https://mikabrv.io/aoa-prediction/aoa_estimation.html))

# Distribución previa: predicciones

# Respuestas observadas



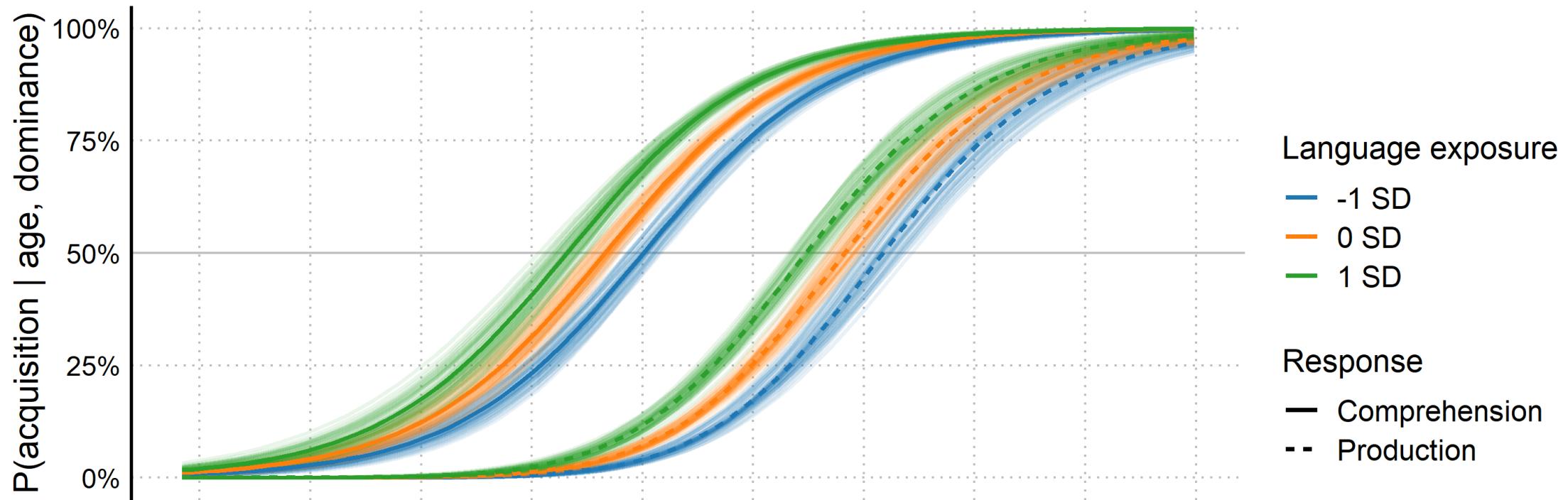
# Distribución posterior



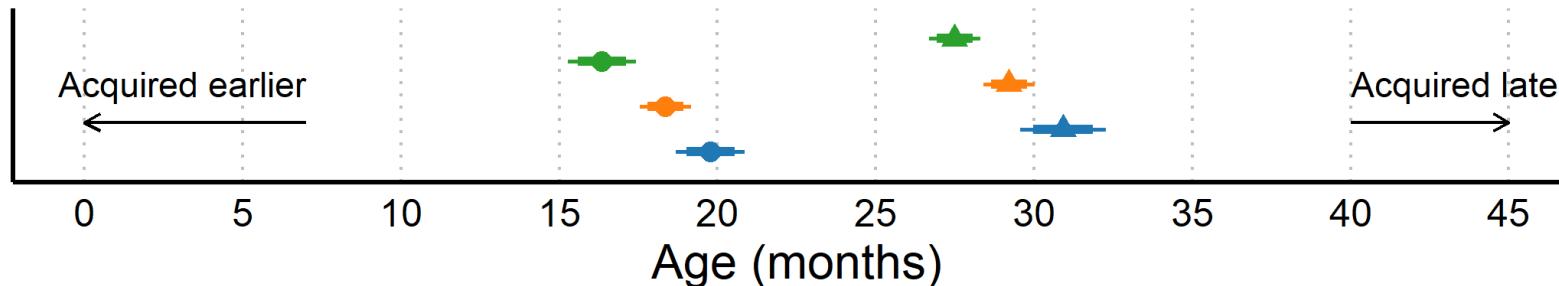
**A**

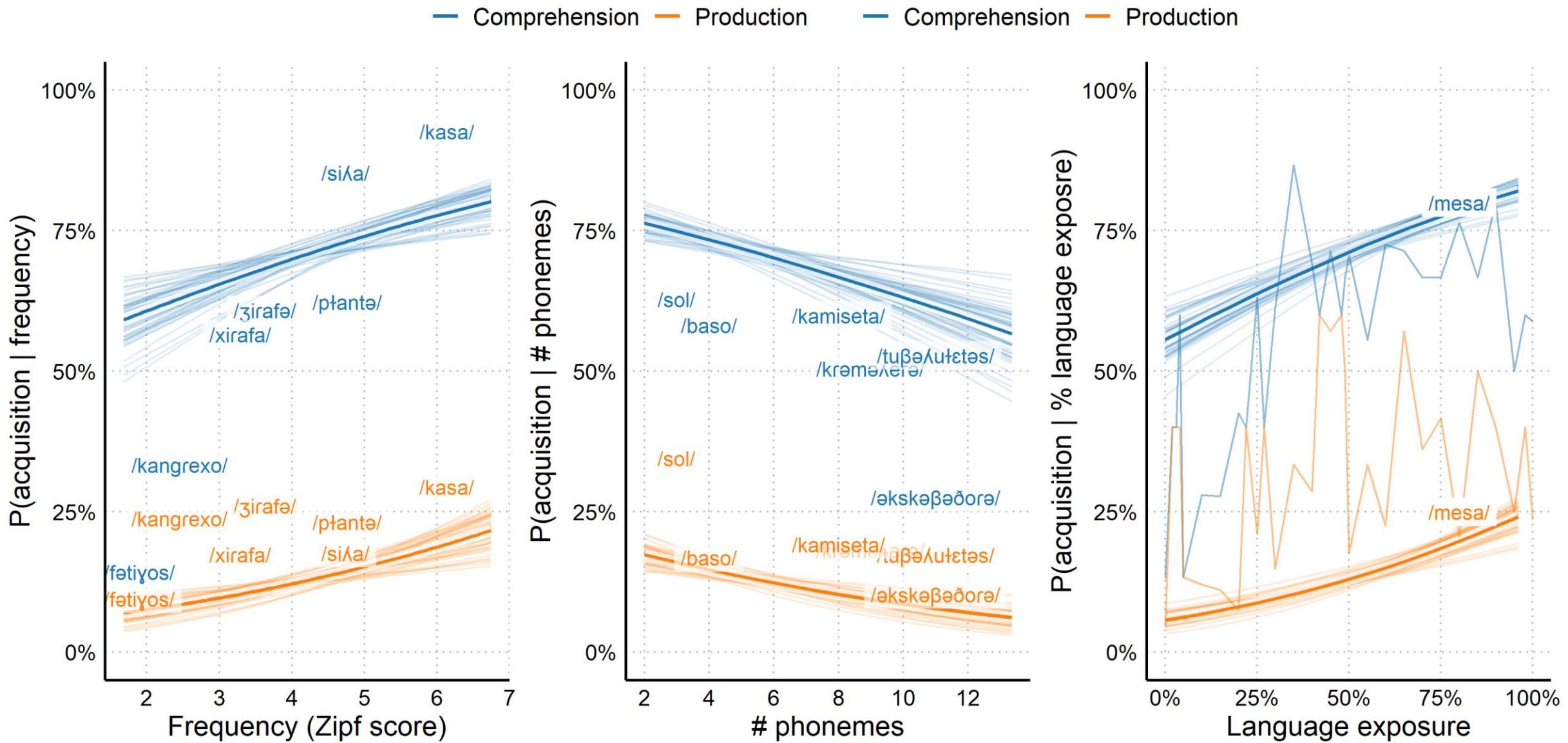
## Acquisition curves

Each line corresponds to a posterior prediction

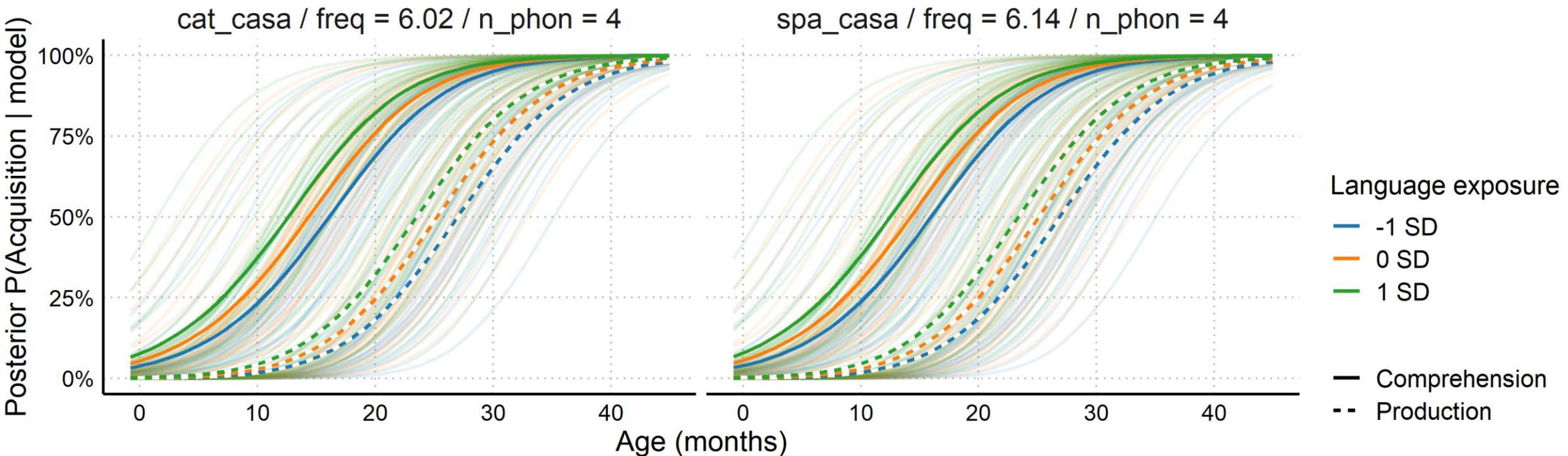
**B**

## Age of acquisition



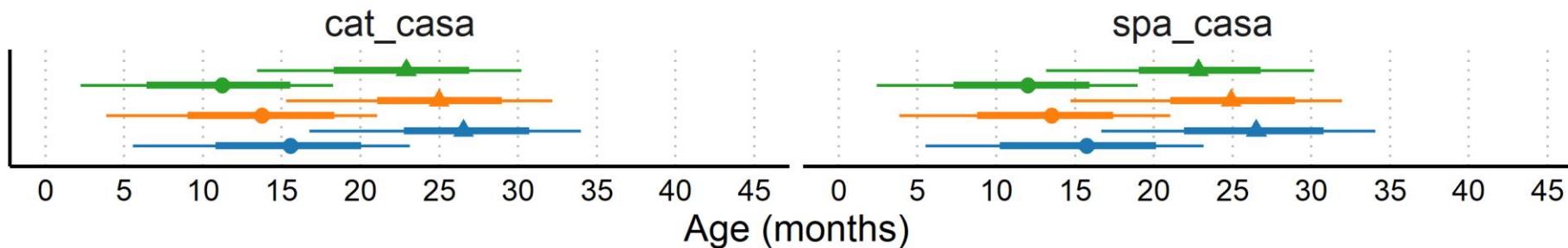


A

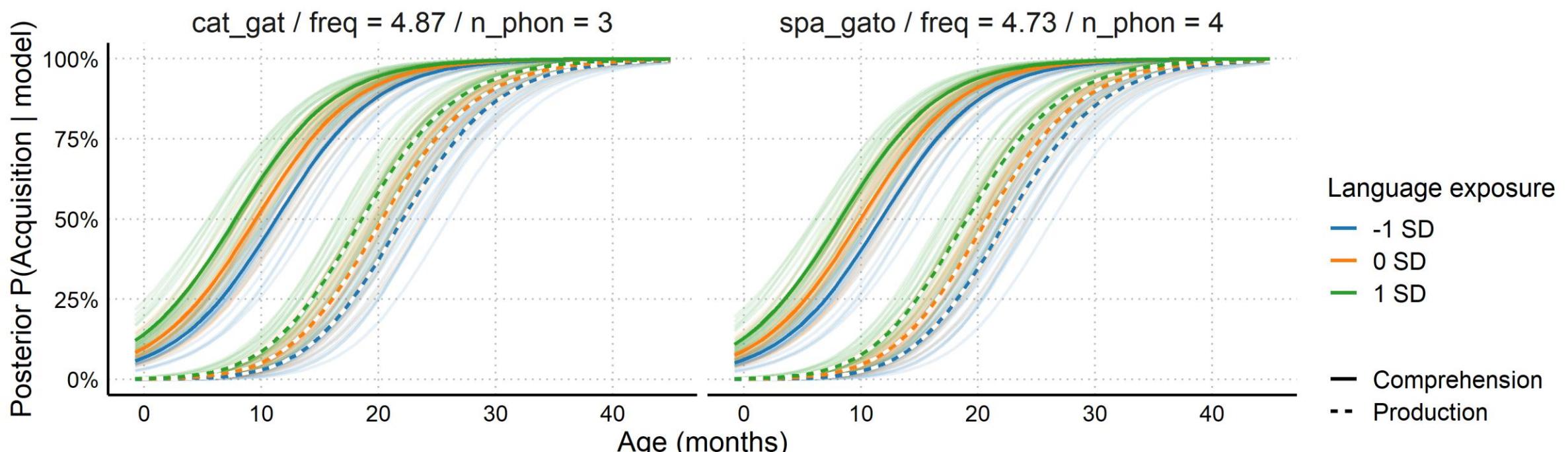


B

### Age of acquisition

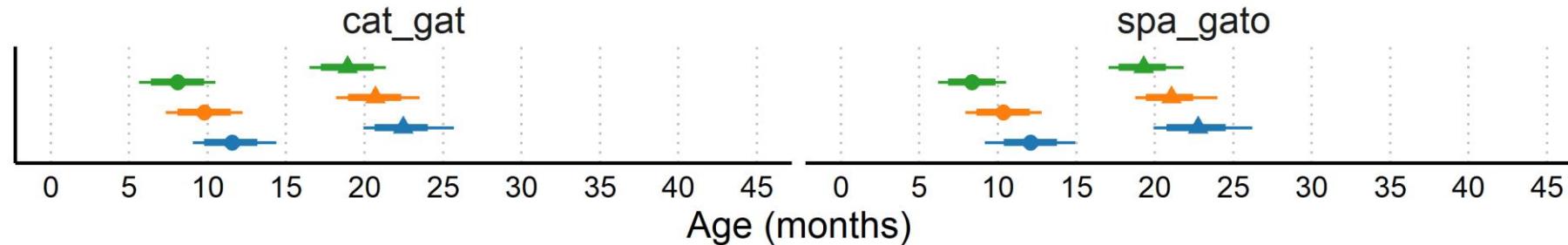


A

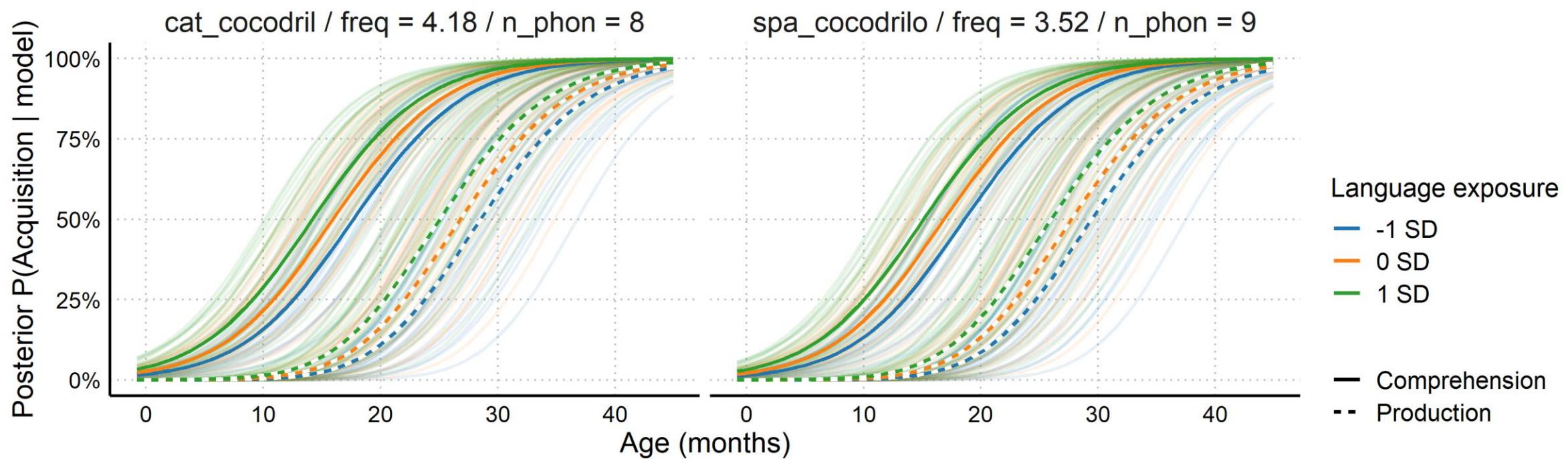


B

### Age of acquisition

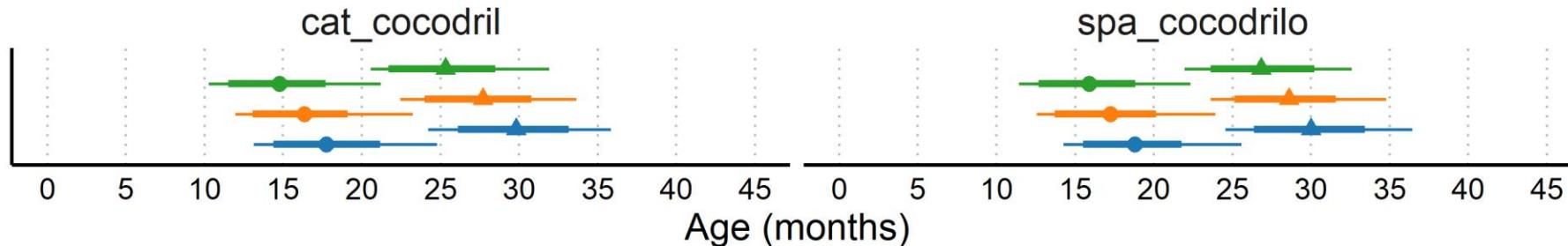


A



B

### Age of acquisition



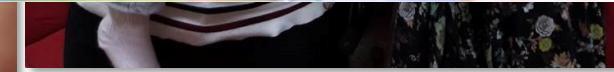
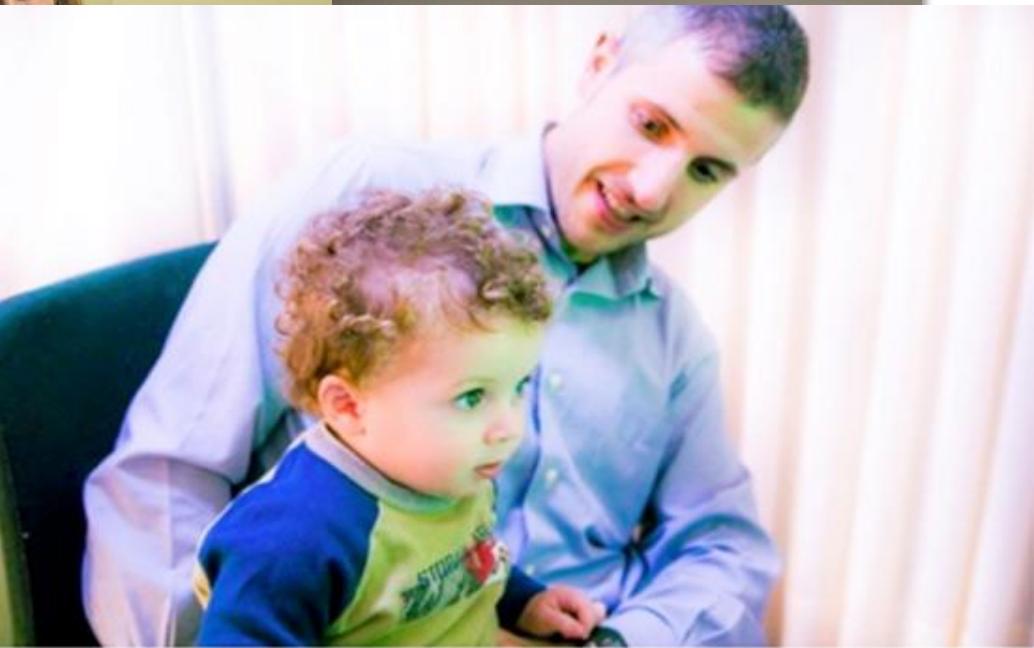
# Conclusiones

Modelo, estimaciones y predicciones útiles

Muestra relativamente limitada (SES)

Futuros pasos:

- Incrementar tamaño de muestra (y representatividad)
- Poner a prueba modelos de crecimiento de vocabulario



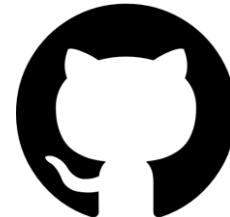


# ¡Gracias!





Código  
Figuras

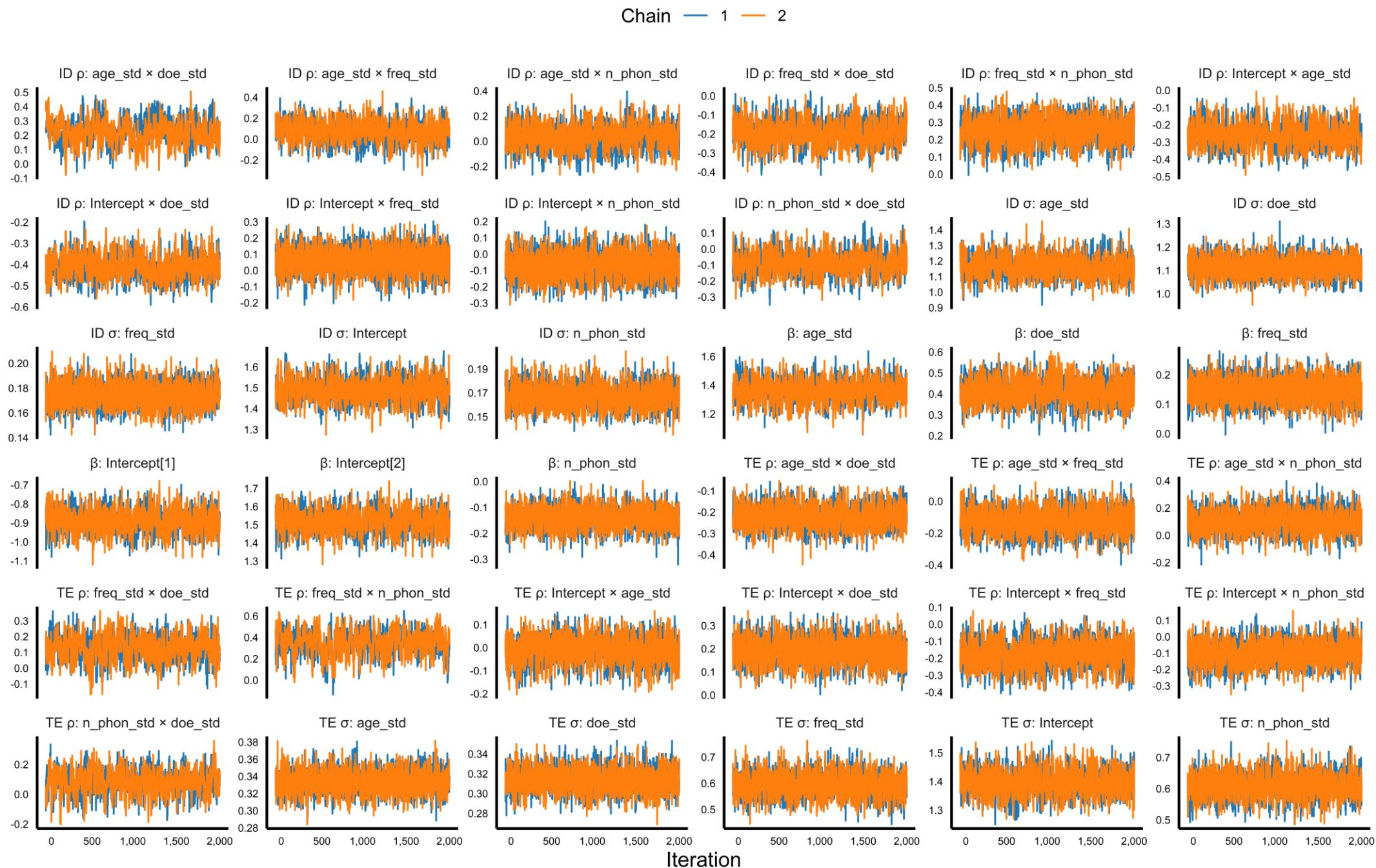


Cosas que he olvidado decir  
Cosas que he ocultado deliberadamente



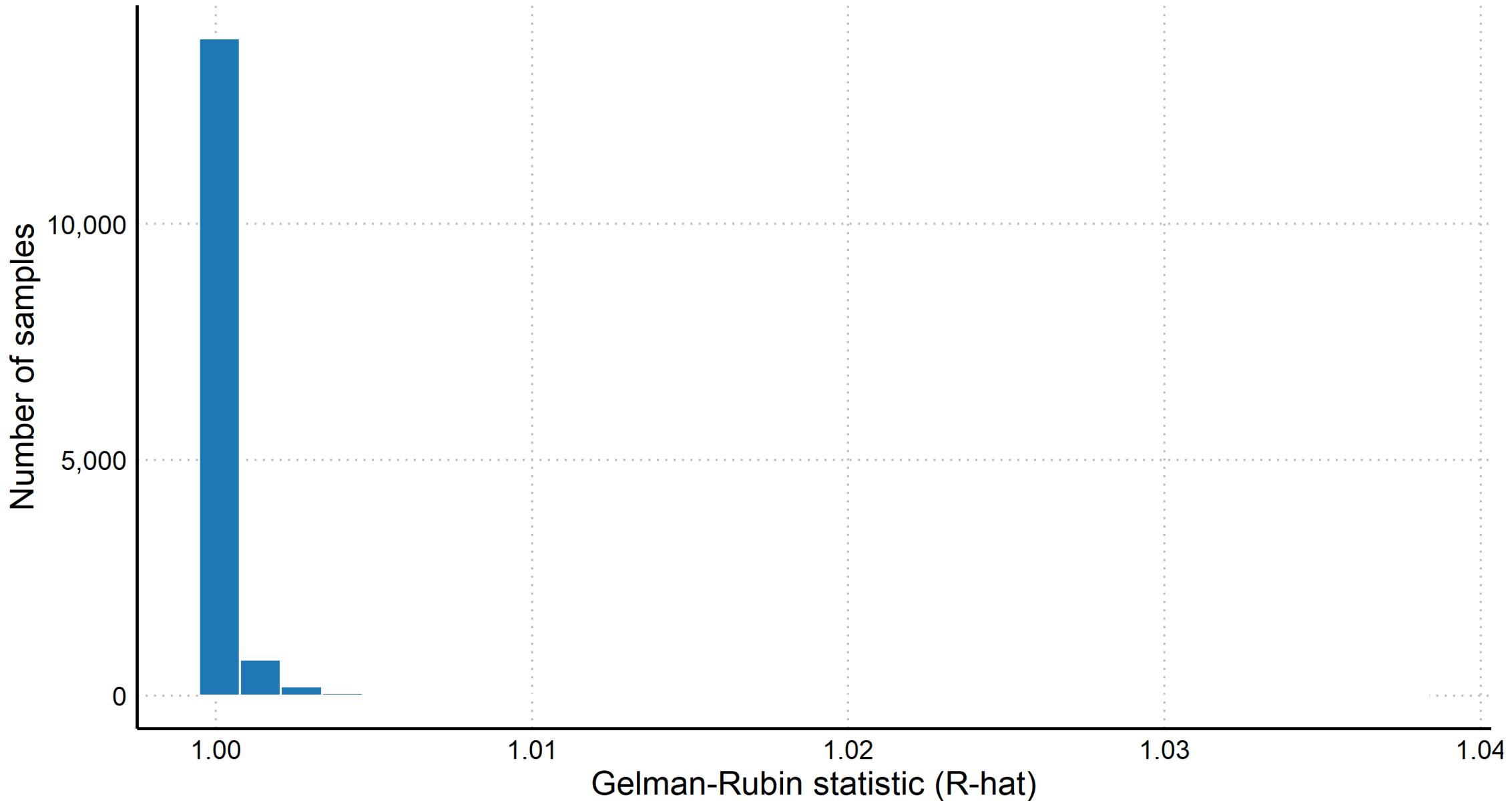
Shiny App





## Chain convergence

Values should not be larger than 1.1



## Effective sample size

Overall, values should be larger than 1

