University of Antwerp
Faculty of Social Sciences

Objective rating method: Entropy

Speech intelligibility estimation

Jose Rivera

April 25, 2022

What are we going to talk about?

1. Preliminars

Research question

Research question

On two fronts:

1. Can comparative judgement (CJ) methods be used to assess speech intelligibility (SI)?,
   To investigate this wee need:
   - an objective measure of SI
2. where CJ stands versus absolute holistic judgement (HJ) methods?,
   In terms of:
   - validity
   - reliability
   - statistical efficiency
   - time efficiency

Objective measure of SI

the most objective (we know of) measure of SI comes from a transcription task:

1. transcribe children's utterances (made by multiple judges),
2. align transcriptions at the utterance level,
3. calculate an entropy measure (H), per utterance per child:

$$H = H(\mathbf{p}) = \frac{-\sum_{i=1}^{n} p_i \cdot \log_2(p_i)}{\log_2(N)}$$

4. characteristics of H [3, 7]
   - bounded in $[0, 1]$ space,
   - utterances with more agreement are more intelligible, and therefore $H \to 0$,
   - utterances with low agreement are less intelligible, and therefore $H \to 1$.

1. Preliminars

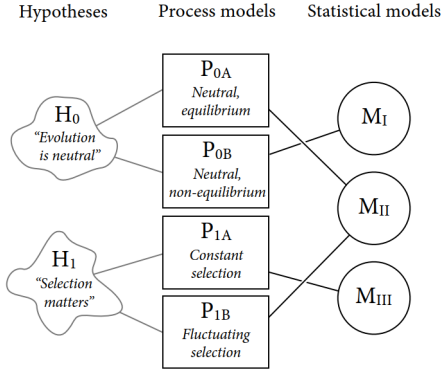
Research hypothesis production

A typical scientific lab

What is needed / deal with[a]

1. Quality of theory
2. Quality of data
3. Reliable procedures and code
4. Quality of data analysis
5. Documentation
6. Reporting

[a] Figure extracted from McElreath [11]



Research hypothesis production[1]

Well known challenges

- Insufficient data
- Wrong population
- Measurement error
- Selection bias
- Confounding

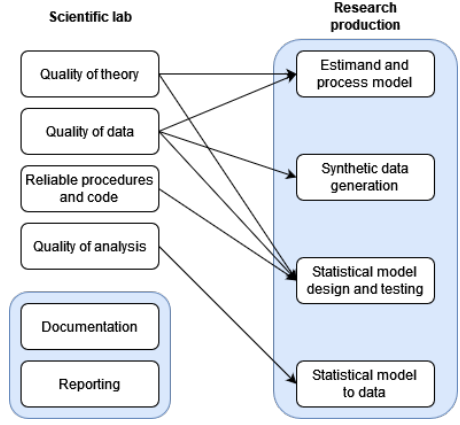
Known challenges in our research;

- Insufficient data (possibly)
- Wrong population
- Measurement error
- Selection bias
- Confounding

[1] Hernán [9], lesson 4

Research hypothesis schematics[2]

a. Estimand and process model
b. Synthetic data generation
c. Statistical model design and testing
d. Apply statistical model to data



[2]McElreath [12], lecture 20, Pearl [13]. Follow Fogarty et al. [8] on item (c).

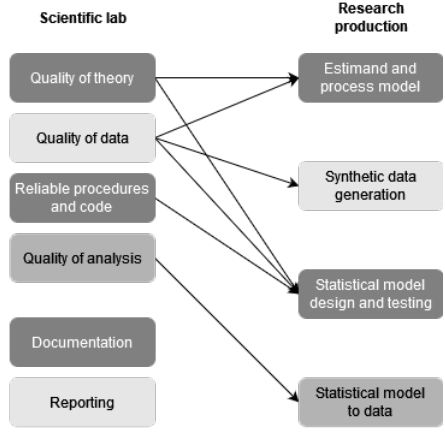
Why do we need to follow this?

Because the improvement of:

- A clear definition of the estimand and process model (assumptions).
- An improved the reliability of your procedures.
- As a documentation procedure.

leads to:

- A sound analysis and results (even when we cannot answer our question).
- An improved planning to get data.

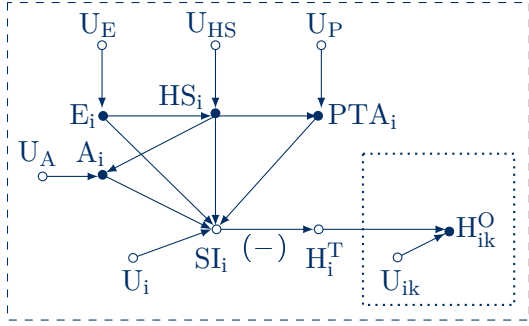


2. Research hypothesis procedure

Estimand and process model

The theory behind our research

- H_{ik} = (observed) entropy replicates
- H_i = (latent) "true" entropy
- SI_i = (latent) SI score
  (inversely related to H_i^T)
- A_i = "hearing" age (minimum)
- E_i = etiology of disease
- HS_i = hearing status
- PTA_i = pure tone average
  (standardized)
- variables assumed independent,
  beyond the described relationships,



General structural diagram

$$P(\mathbf{U}) = P(U_{ik}, U_i, U_A, U_E, U_{HS}, U_P)$$
$$= P(U_{ik})P(U_i)P(U_A)P(U_E)P(U_{HS})P(U_P)$$

Probabilistic (causal) model

First form

$$
\begin{aligned}
\mathrm{H}^{\mathrm{O}}_{\mathrm{ik}} &\leftarrow \mathrm{f}(\mathrm{H}^{\mathrm{T}}_{\mathrm{i}}, \mathrm{U}_{\mathrm{ik}}) \\
\mathrm{H}^{\mathrm{T}}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{SI}_{\mathrm{i}}) \\
\mathrm{SI}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{HS}_{\mathrm{i}}, \mathrm{E}_{\mathrm{i}}, \mathrm{A}_{\mathrm{i}}, \mathrm{PTA}_{\mathrm{i}}, \mathrm{U}_{\mathrm{i}}) \\
\\
\\
\\
\mathrm{HS}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{HS}}) \\
\mathrm{A}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{A}}) \\
\mathrm{E}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{E}}) \\
\mathrm{PTA}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{P}}) \\
\mathrm{U} &\sim \mathrm{P}(\mathbf{U})
\end{aligned}
$$

(a) general structural model

$$
\begin{aligned}
\mathrm{H}^{\mathrm{O}}_{\mathrm{ik}} &\sim \mathrm{BetaProp}(\mathrm{H}^{\mathrm{T}}_{\mathrm{i}}, \mathrm{M}_{\mathrm{ik}}) \\
\mathrm{H}^{\mathrm{T}}_{\mathrm{i}} &= \mathrm{inv\_logit}(-\mathrm{SI}_{\mathrm{i}}) \\
\mathrm{SI}_{\mathrm{i}} &\sim \mathrm{Normal}(\mu_{\mathrm{SI}}, \sigma_{\mathrm{Ui}}+) \\
\mu_{\mathrm{SI}} &= \alpha + \alpha_{\mathrm{E[i],HS[i]}} + \beta_{\mathrm{P}}\mathrm{PTA}_{\mathrm{i}} \\
&\quad + \beta_{\mathrm{A,HS[i]}}(\mathrm{A}_{\mathrm{i}} - \bar{\mathrm{A}}) \\
\mathrm{HS}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{A}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{E}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{PTA}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{U} &\sim \mathrm{unobservable}
\end{aligned}
$$

(a) general probabilistic model

Probabilistic (causal) model

First form

Notice,

- α, $\alpha_{\text{HS[i]}}$, $\alpha_{\text{E[i]}}$, $\beta_{\text{A,HS[i]}}$, β_{P} are structural parameters (as in SEM)
- U_{ik} = replicates measurement error U_{i} = between child SI variability
- variability of U_{ik} is modeled by M_{ik} variability of U_{i} is modeled by σ_{Ui}

$$
\begin{aligned}
\text{H}^{\text{O}}_{\text{ik}} &\sim \text{BetaProp}(\text{H}^{\text{T}}_{\text{i}}, \text{M}_{\text{ik}}) \\
\text{H}^{\text{T}}_{\text{i}} &= \text{inv\_logit}(-\text{SI}_{\text{i}}) \\
\text{SI}_{\text{i}} &\sim \text{Normal}(\mu_{\text{SI}}, \sigma_{\text{Ui}}) \\
\mu_{\text{SI}} &= \alpha + \alpha_{\text{E[i],HS[i]}} + \beta_{\text{P}}\text{PTA}_{\text{i}} \\
&\quad + \beta_{\text{A,HS[i]}}(\text{A}_{\text{i}} - \bar{\text{A}}) \\
\text{HS}_{\text{i}} &\sim \text{data} \\
\text{A}_{\text{i}} &\sim \text{data} \\
\text{E}_{\text{i}} &\sim \text{data} \\
\text{PTA}_{\text{i}} &\sim \text{data} \\
\text{U} &\sim \text{unobservable}
\end{aligned}
$$

(a) general probabilistic model

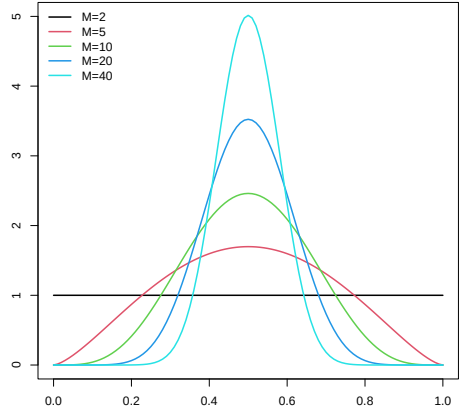
Probabilistic (causal) model

Express variability in BetaProp

$$
\begin{aligned}
H^{O}_{ik} &\sim \text{BetaProp}(H^{T}_{i}, M_{ik}) \\
H^{T}_{i} &= \alpha/(\alpha+\beta) \\
M_{ik} &= \alpha+\beta
\end{aligned}
$$

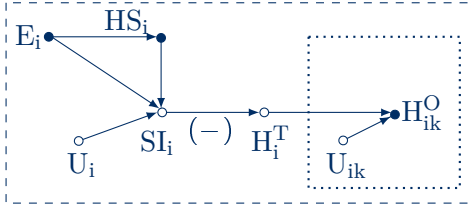
$$
\begin{aligned}
\alpha &= H^{T}_{i} \cdot M_{ik} \\
\beta &= (1-H^{T}_{i}) \cdot M_{ik}
\end{aligned}
$$

$$
\begin{aligned}
\alpha &= 0.5 \cdot 2 \\
\beta &= (1-0.5) \cdot 2
\end{aligned}
$$

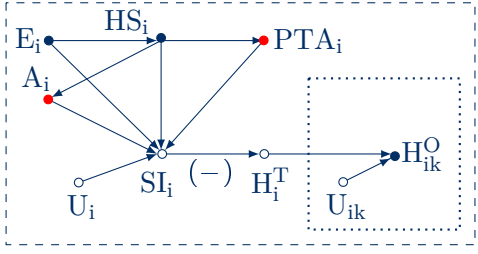


Interested in two effects

1. total effects model inherits:
   - children's characteristics that lead to the fitting of specific apparatus,
   - the (convenience of) sample selection (fixed with post-stratification)
2. to do the last, we stratify for all variables that explain variability, ergo, use a direct effects model
3. two levels: replicates (k), children (i), denoted by discontinuous squares



(b) total effects



(a) direct effects

Probabilistic (causal) model

Second form

$$
\begin{aligned}
\mathrm{H}^{\mathrm{O}}_{\mathrm{ik}} &\leftarrow \mathrm{f}(\mathrm{H}^{\mathrm{T}}_{\mathrm{i}}, \mathrm{U}_{\mathrm{ik}}) \\
\mathrm{H}^{\mathrm{T}}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{SI}_{\mathrm{i}}) \\
\mathrm{SI}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{HS}_{\mathrm{i}}, \mathrm{E}_{\mathrm{i}}, \mathrm{A}_{\mathrm{i}}, \mathrm{PTA}_{\mathrm{i}}, \mathrm{U}_{\mathrm{i}}) \\
\\
\mathrm{HS}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{HS}}) \\
\mathrm{A}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{A}}) \\
\mathrm{E}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{E}}) \\
\mathrm{PTA}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{P}}) \\
\mathrm{U} &\sim \mathrm{P}(\mathbf{U})
\end{aligned}
$$

(a) general structural model

$$
\begin{aligned}
\mathrm{H}^{\mathrm{O}}_{\mathrm{ik}} &\sim \mathrm{BetaProp}(\mathrm{H}^{\mathrm{T}}_{\mathrm{i}}, \mathrm{M}_{\mathrm{ik}}) \\
\mathrm{H}^{\mathrm{T}}_{\mathrm{i}} &= \mathrm{inv\_logit}(-\mathrm{SI}_{\mathrm{i}}) \\
\mathrm{SI}_{\mathrm{i}} &= \mathrm{a}_{\mathrm{i}} + \alpha + \alpha_{\mathrm{E[i],HS[i]}} + \beta_{\mathrm{P}}\mathrm{PTA}_{\mathrm{i}} \\
&\quad + \beta_{\mathrm{A,HS[i]}}(\mathrm{A}_{\mathrm{i}} - \bar{\mathrm{A}}) \\
\mathrm{a}_{\mathrm{i}} &\sim \mathrm{Normal}(0, \sigma_{\mathrm{Ui}}) \\
\mathrm{HS}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{A}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{E}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{PTA}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{U} &\sim \mathrm{unobservable}
\end{aligned}
$$

(a) general probabilistic model

2. Research hypothesis procedure

Synthetic data generation

Idealized data[3]

Simulation data can serve as [10, 11],

1. A place where to test your model, on multiple purposes,
   - parameter recovery
   - power
2. A reflection of a population,
   - De Raeve [6]: 70 HI/CI, 130 HI/HA
   - Our idealized data: 150 NH, 70 HI/CI, 130 HI/HA
3. A reflection of a hypothesis,
   - size of effects

```
(sim_name=NULL, # file_name need to includ
sim_save=NULL, # file_save need to includ
seed=NULL, # seed
I=350, # experimental units (children)
K=10, # replicates (utterances)
p=c(0.50, 0.175, 0.325), # children prop.
par=list( m_i=0, s_i=0.5, # hyperprior ch
          m_M=10, s_M=NULL, # generation
          a=0,
          bP=-0.1,
          aHS=c(0.4,0,-0.4),
          bA=0.15,
          bAHS=rep(0,3),
          aE=rep(0,4),
          aEHS=matrix( c( rep(0,4), # NH
                          c(0, seq(0.1,-0
                          c(0, seq(0.1,-0
                        ncol=3, byrow=F) )
```

[3] more details in file: 1_2_E_sim_fun.R

Idealized data

About the size of the effects
(in logits, no previous info)

1. E has a full interaction with HS
2. aHS[j] − aHS[i] ≈ −0.4,
   NH vs HI/CI (depends on E)
3. bP = −0.1, per PTA unit
   (+10 PTA units ⇒ −1 logit),
4. bA ≈ 0.15, per A unit, above the
   minimum (depends on HS)
   (+10 A units ⇒ +1.5 logits)

```
(sim_name=NULL, # file_name need to includ
sim_save=NULL, # file_save need to includ
seed=NULL, # seed
I=350, # experimental units (children)
K=10, # replicates (utterances)
p=c(0.50, 0.175, 0.325), # children prop.
par=list( m_i=0, s_i=0.5, # hyperprior ch
          m_M=10, s_M=NULL, # generation
          a=0,
          bP=-0.1,
          aHS=c(0.4,0,-0.4),
          bA=0.15,
          bAHS=rep(0,3),
          aE=rep(0,4),
          aEHS=matrix( c( rep(0,4), # NH
                          c(0, seq(0.1,-0
                          c(0, seq(0.1,-0
                       ncol=3, byrow=F) )
```

Idealized data

1. variables are generated in a random fashion
2. between children SI variability are defined by the random effects

```r
# 1. true data ####
dT = data.frame(matrix(NA, nrow=I, ncol=1))
names(dT) = c('child_id')
dT$child_id = 1:I


# assigning children to groups
n = round( p*I )
if( sum(n) != I ){
  if( I - sum(n[c(1,3)]) > n[2] ){
    n[2] = I - sum(n[c(1,3)]) # to sum the right amount
  } else {
    n[3] = I - sum(n[c(1,3)]) # to sum the right amount
  }
}


# generating covariates
if(!is.null(seed)){
  set.seed(seed+1)
}
dT$HS = c( rep(1, n[1]), rep(2, n[2]), rep(3, n[3]))
dT$A = round( rnorm( sum(n), 5, 1) )
dT$A = with(dT, ifelse(A>7, 7, A) )

dT$E = c( rep(1, n[1]), # no way to know true effects
          sample(2:3, size=n[2], replace=T),
          sample(3:4, size=n[3], replace=T))

dT$PTA = c( round(rnorm(n[1], 60, 15)), # first 12 NH
            round(rnorm(n[2], 90, 15)), # next 10
            round(rnorm(n[3], 110, 15))) # last 10


# children's random effects
if(!is.null(seed)){
  set.seed(seed-1)
}
par$re_i = rnorm(I, par$m_i, par$s_i)
dT$re_i = par$re_i # children's random effects (between SI v
```

Idealized data

1. we use second form of the probabilistic model
2. “true’ entropy (Ht) is inversely related to SI
3. we simulate measurement error through M from BetaProp() distribution.

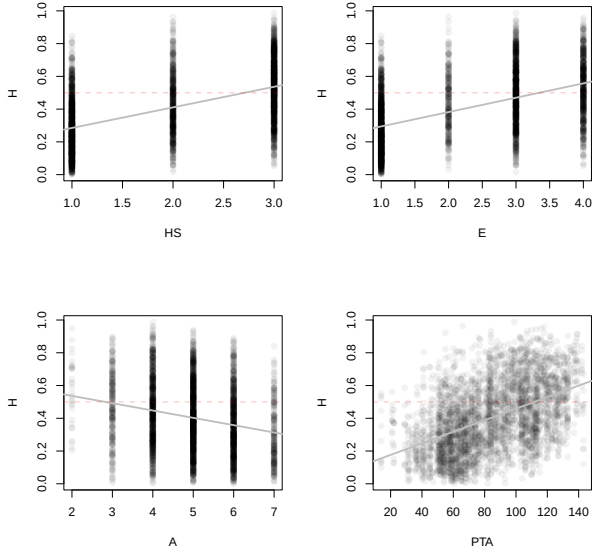
```
# linear predictor / SI index
dT$SI = NA
A_bar = min(dT$A)
sPTA = standardize( dT$PTA )
for(i in 1:I){
  dT$SI[i] = with(dT, par$re_i[i] +
                      par$a +
                      par$bP*sPTA[i] +
                      par$aHS[ HS[i] ] +
                      par$aE[ E[i] ] +
                      par$aEHS[ E[i], HS[i] ] +
                      par$bA*( A[i] - A_bar ) +
                      par$bAHS[ HS[i] ]*( A[i] - A_bar ) )
}


# true entropy
dT$Ht = inv_logit(-dT$SI) # true entropy (SI -> Ht: negati


# variability of H
if(!is.null(seed)){
  set.seed(seed+2)
}
if( is.numeric(par$m_M) & !is.numeric(par$s_M) ){
  par$M = rep(par$m_M, I)
} else{
  par$M = round( rlnorm(I, meanlog=par$m_M, sdlog= par$s_M
}
dT$M = par$M # same df for all children (not same shape!!)


# rounding
dT[,6:ncol(dT)] = round( dT[,6:ncol(dT)], 5)
```

Idealized data

1. we simulate replicate measures of entropy (H)
2. we storage all relevant parameters and data

```
# 2. observed data ####
N = I*K
dO = data.frame(matrix(NA, nrow=N, ncol=3))
names(dO) = c('child_id','utt_id','H')
dO$child_id = rep(1:I, each=K)
dO$utt_id = rep(1:K, I)

# generating observed H
# i=1
if(!is.null(seed)){
  set.seed(seed-2)
}
for(i in 1:I){

  # identify data
  idx = dO$child_id == i

  # linear predictor
  dO$H[idx] = rbeta2(n=K, prob=dT$Ht[i], theta=dT$M[
}

# round
dO$H = round( dO$H, 5)
```

Example



2. Research hypothesis procedure

Statistical model design and testing

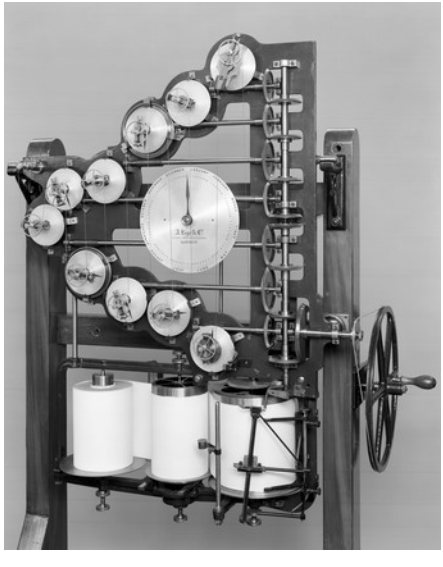
Model design and test[4]

Purpose:

- to have reliable procedures,
- to maintain a clear documentation,
- to have a sound analysis

Procedure:

- step by step, instantiating one difficulty at the time
- use probabilistic assumptions defined in estimand and process model
- is like running synthetic data generation backwards

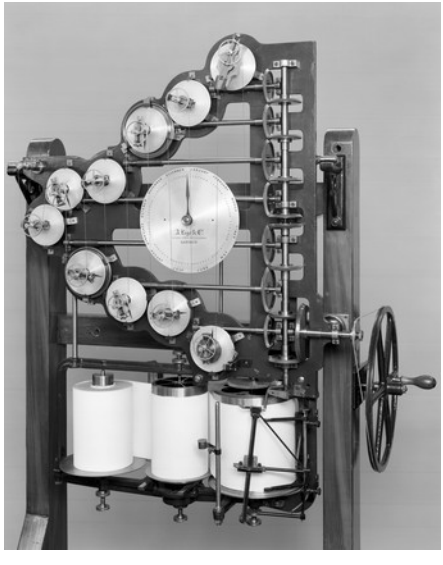


[4] Following Fogarty et al. [8]

Model design and test

We evaluate,

- the probabilistic model implementation [11, 2] (centered and non-centered versions)
- prior predictive
- "health" of MCMC chains
- parameter recovery
- posterior predictive
- power



Probabilistic model

starts with,

- the simplest model
- the simplest data generating procedure

```
transformed parameters{
    vector[I] SI;          // SI index (per child)
    vector[I] Ht;          // true entropy (per chi

    SI = a + re_i;         // linear predictor
    Ht = inv_logit(-SI);   // average entropy (SI -
}
model{

    // hyperpriors
    m_i ~ normal( 0 , 0.2 );
    s_i ~ exponential( 1 );

    // priors
    a ~ normal( 0 , 0.2 );
    re_i ~ normal( m_i , s_i );

    // likelihood
    for(n in 1:N){
      H[n] ~ beta_proportion( Ht[cid[n]] , 10 );
    }
}
```

Probabilistic model

try,

- the centered and non-centered parametrization

```
transformed parameters{
    vector[I] re_i;       // random intercepts (per
    vector[I] SI;         // SI index
    vector[I] Ht;         // true entropy (per child

    re_i = m_i + s_i*z_re;// non-centered RE
    SI = a + re_i;        // linear predictor
    Ht = inv_logit(-SI);  // average entropy (SI ->
}
model{

    // hyperpriors
    m_i ~ normal( 0 , 0.2 );
    s_i ~ exponential( 1 );

    // priors
    a ~ normal( 0 , 0.2 );
    z_re ~ std_normal();

    // likelihood
    for(n in 1:N){
      H[n] ~ beta_proportion( Ht[cid[n]] , 10 );
    }
}
```

Probabilistic model

escalate,

- complexity of model
- traits of the data

```
transformed parameters{
    vector[I] SI;          // SI index (per child)
    vector[I] Ht;          // true entropy (per chil

    // linear predictor
    for(i in 1:I){
      SI[i] = re_i[i] + a + aHS[HS[i]] +
        bA*Am[i] + bP*sPTA[i];
      // no multicollinearity between E and HS
    }

    // average entropy (SI -> Ht: negative)
    Ht = inv_logit(-SI);  // average entropy (SI ->
}
model{

    // hyperpriors
    m_i ~ normal( 0 , 0.2 );
    s_i ~ exponential( 1 );

    // priors
    a ~ normal( 0 , 0.2 );
    re_i ~ normal( m_i , s_i );
    //aE ~ normal( 0 , 0.5 );
    aHS ~ normal( 0 , 0.5 );
    bP ~ normal( 0 , 0.3 );
    bA ~ normal( 0 , 0.3 );
    m_M ~ lognormal( 1.5 , 0.5 );

    // likelihood
    for(n in 1:N){
      H[n] ~ beta_proportion( Ht[cid[n]] , m_M );
    }
}
```

Probabilistic model

We tested 5 random effects models:
(from 5 synthetic data types)
(centered, and non-centered)

- only intercept, M = 10,
- causal model, M = 10,
- causal model, M per individual,
- no known process,
- causal model with interactions, M per individual,

```
transformed parameters{
    vector[I] SI;          // SI index (per child)
    vector[I] Ht;          // true entropy (per child)

    // linear predictor
    for(i in 1:I){
      SI[i] = re_i[i] + a + aHS[HS[i]] +
        bA*Am[i] + bP*sPTA[i];
      // no multicollinearity between E and HS
    }

    // average entropy (SI -> Ht: negative)
    Ht = inv_logit(-SI);  // average entropy (SI -> H
}
model{

    // hyperpriors
    m_i ~ normal( 0 , 0.2 );
    s_i ~ exponential( 1 );

    // priors
    a ~ normal( 0 , 0.2 );
    re_i ~ normal( m_i , s_i );
    //aE ~ normal( 0 , 0.5 );
    aHS ~ normal( 0 , 0.5 );
    bP ~ normal( 0 , 0.3 );
    bA ~ normal( 0 , 0.3 );
    m_M ~ lognormal( 1.5 , 0.5 );

    // likelihood
    for(n in 1:N){
      H[n] ~ beta_proportion( Ht[cid[n]] , m_M );
    }
}
```

Prior predictive

Priors and hyper-priors

- In the probabilistic (causal) model there were no priors for our parameters,
- To decide our priors we follow McElreath [11]: "priors are part of the assumptions, and should be inspected as such",
- We will evaluate the implications of our priors on the outcome scale. We have three outcomes scales: SI_i, H_i^T, and H_{ik}^O

Priors

$$
\begin{aligned}
a_i &\sim \text{Normal}(\mu_a, \sigma_a) \\
M_i &\sim \text{LogNormal}(\mu_M, \sigma_M) \\
\alpha &\sim \text{Normal}(0, 0.2) \\
\alpha_{HS[i]} &\sim \text{Normal}(0, 0.3) \\
\alpha_{E[i]} &\sim \text{Normal}(0, 0.3) \\
\beta_{A,HS[i]} &\sim \text{Normal}(0, 0.3) \\
\beta_P &\sim \text{Normal}(0, 0.3)
\end{aligned}
$$

Hyper-priors

$$
\begin{aligned}
\mu_a &\sim \text{Normal}(0, 0.2) \\
\sigma_a &\sim \text{Exp}(1) \\
\mu_M &\sim \text{Normal}(0, 0.5) \\
\sigma_M &\sim \text{Exp}(1)
\end{aligned}
$$

Prior predictive

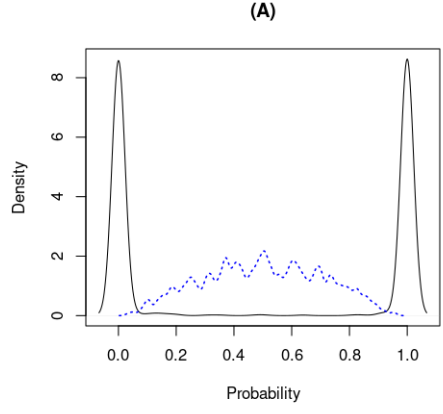
Undesired assumptions can easily creep in non-linear models[a]
Example:

(black line)

$$\theta \sim \mathrm{N}(0, 100)$$
$$\mathrm{logit}(\mathrm{p}) = \theta$$

(blue line)

$$\theta \sim \mathrm{N}(0, 1)$$
$$\mathrm{logit}(\mathrm{p}) = \theta$$



[a] Figure extracted from Rivera [14].

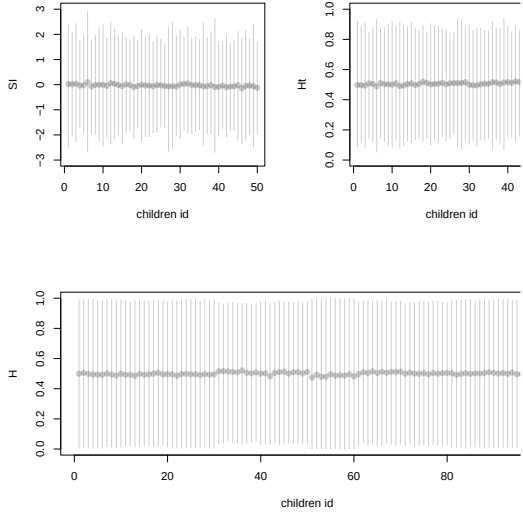
Prior predictive

What our priors imply?

NO undesired assumption has crept in:

- the SI_i scale,
- the H_i^T scale,
- the H_{ik}^O scale

i.e. the scales' full space can be reached by (a combination of) the parameters

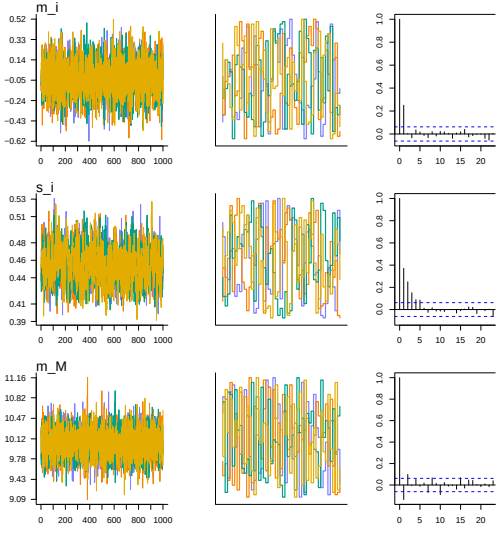


“Health” of MCMC chains

The MCMC chains achieve,

- good convergence
- good mixing
- lack of autocorrelation

same results on the n_eff and RHat statistics.

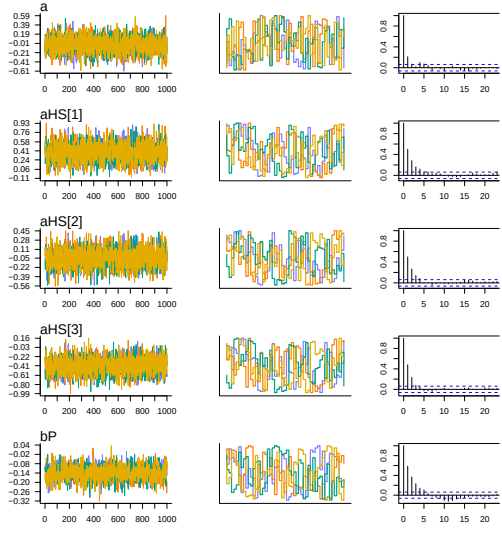


“Health” of MCMC chains

The MCMC chains achieve,

- good convergence
- good mixing
- lack of autocorrelation

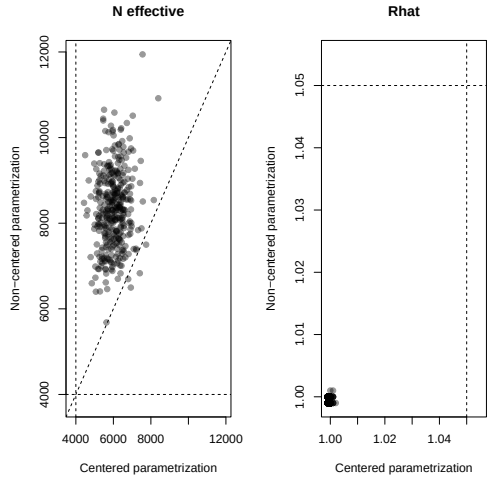
same results on the n_eff and RHat statistics.



“Health” of MCMC chains

The non-centered parametrization has,

- better n_eff
  (denoting lack of autocorrelation)
- better Rhat
  (denoting good convergence)
- better mixing
  (inspected visually, not shown)



Parameter recovery

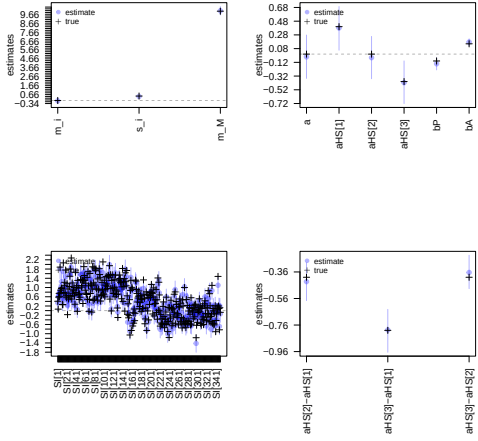
On idealized data

The model,

with less working assumptions[a]

- recovers the parameters in the right scale,
- most of the "true" parameters are inside of the compatibility intervals (CI)
- contrast are approximately correct

[a] random effects causal model, with M = 10



Posterior predictive

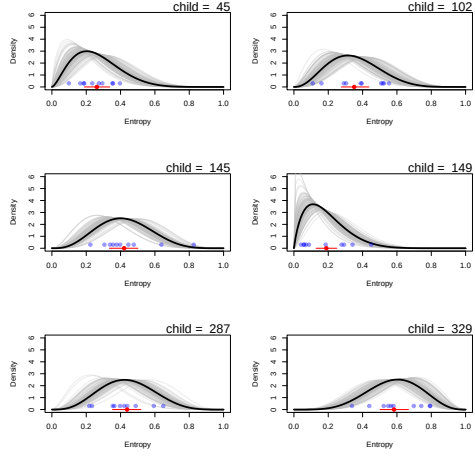
But how well reproduces the data?,
with less working assumptions[a]

- captures the variability of the replicates,
- provides a "true" H and SI

More complex models,

- captures even better the data, but might overfit,
- avoid overfit when the model is fit to the data (ITA [1, 4])

[a] random effects causal model, with M = 10



Power

Equivalent prior sampling method [18, 10]

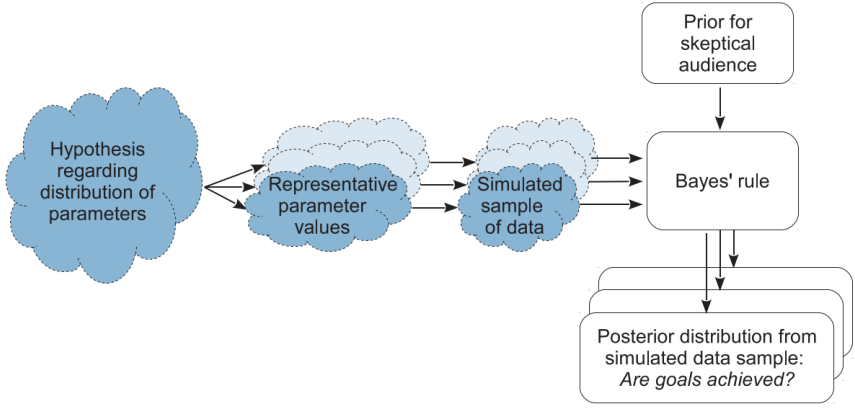
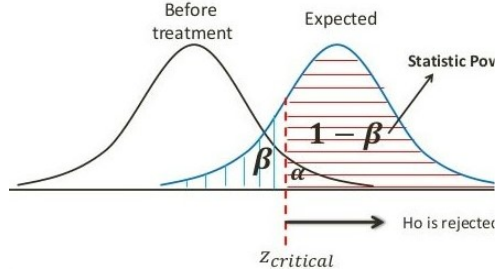


Figure: Flow of information in a power analysis, in which simulated data come from random hypothetical parameters. Extracted from Kruschke [10].

Power

Equivalent prior sampling method

1. generate idealized data
2. generate parameters' distribution (with Bayes theorem)
3. simulate data sample (variability in data and parameters)
4. apply models to simulated data (include priors)
5. evaluate desire goals
   - reject the null hypothesis
   - affirm predicted value
   - achieve precision in estimate
6. repeat procedure (to approximate power)



Power

Equivalent prior sampling method

1. ~~generate idealized data~~
2. ~~generate parameters' distribution~~ (with Bayes theorem)
3. simulate data sample (variability in data and parameters)
4. apply models to simulated data (include priors)
5. evaluate desire goals
   - reject the null hypothesis
   - affirm predicted value
   - achieve precision in estimate
6. repeat procedure (to approximate power)

```
Epower( power_save=file.path(getwd(), 'sim_chain'), # po
        sim_name='Hbeta_sim2_power.RData', # file_save n
        sim_save=file.path(getwd(), 'sim_data'), # file_
        model_name='Hbeta_NC_sim2', # model for which we
        model_in=file.path(getwd(), 'sim_models'), # loc
        model_out=file.path(getwd(), 'sim_chain'), # loc
        Nsim=100, # number of simulation for power
        I_grid = c(48, 60), # experimental units (childr
        K_grid = c(10, 20), # replicates (utterances)
        p=C(0.34, 0.33, 0.33), # children prop. on each
        par_int=c('aHS','bP','bA','m_i','s_i','m_M','SI'
        par_cont=c('aHS','SI') ) # parameters to contras
```

Power

Equivalent prior sampling method

1. ~~generate idealized data~~
2. ~~generate parameters' distribution~~ (with Bayes theorem)
3. ~~simulate data sample~~ (variability in data and parameters)
4. ~~apply models to simulated data~~ (include priors)
5. evaluate desire goals[a]
   - reject the null hypothesis
   - affirm predicted value
   - achieve precision in estimate
6. repeat procedure (to approximate power)

[a] use ROPE [10], and effect sizes [5, 15]

```
(true_par,
cuts=c(0,0.1,0.2,0.5,0.8,1.2,2), # based
rvalues=c(0.2,0.2,0.15,0.3,0.3,0.4,0.8))
```

Effect size	*d*
Very small	0.01
Small	0.20
Medium	0.50
Large	0.80
Very large	1.20
Huge	2.0

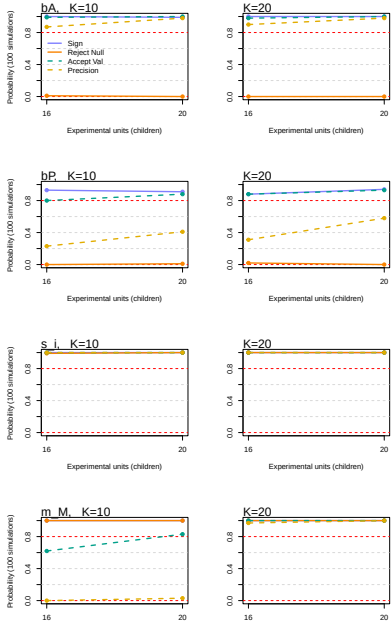
Power results

For this parameter set,

- reject the null hypothesis, never achieved in structural parameters,
- affirm predicted value, in all shown parameters
- achieve precision in estimate, in some parameters is reached

Notice,

- between child SI variability,
- replicates measurement error (M),
- not much difference with 2x comparisons (except for M)



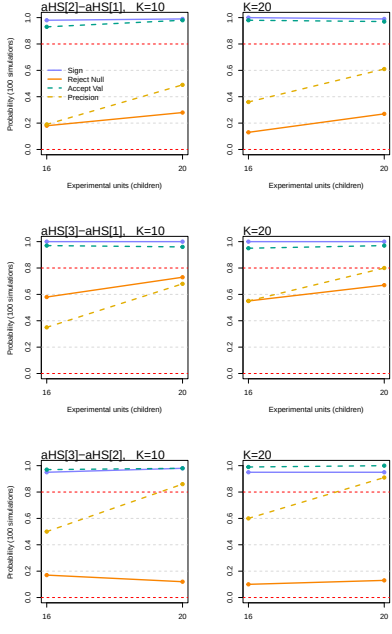
Power results

For group contrasts,

- reject the null hypothesis, closest for the largest contrast,
- affirm predicted value, in all shown contrasts,
- achieve precision in estimate, grows with more children

Notice,

- not much difference with 2x comparisons
- group contrasts are easier to identify (larger sample size per comparison)



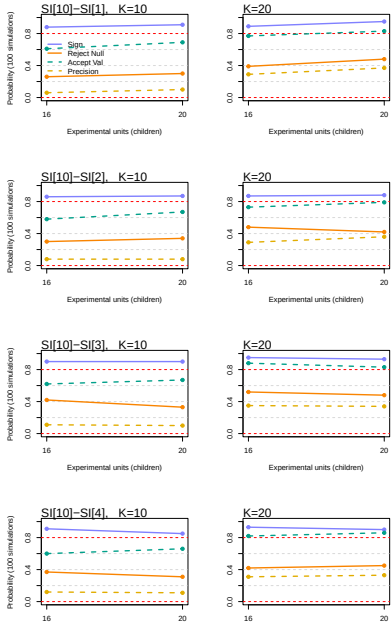
Power results

For individual contrasts,

- reject the null hypothesis, depends on the contrast of interest,
- affirm predicted value, for some is achieved,
- achieve precision in estimate, clearly requires more sample size

Notice,

- we see a clear difference with 2x comparisons (more sample at the appropriate level)

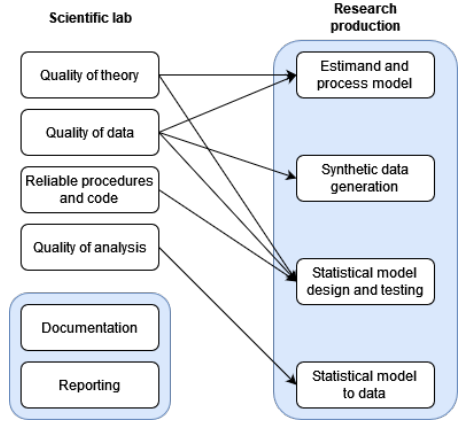


2. Research hypothesis procedure

Apply statistical model to data

What we have so far

1. measure:
   replicated entropies H_{ik}^{O}
2. estimand:
   SI index, structural parameters, contrasts
3. structural models:
   total and direct effects
4. probabilistic models:
   three possible fitting models
5. statistical models:
   works as intended
6. power:
   enough



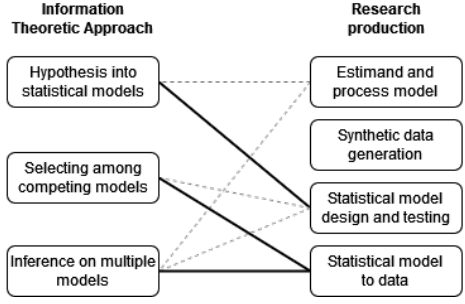
Information Theoretic Approach [1, 4]

The last step would be select the most fitting model using the ITA,

1. ~~hypothesis into statistical models~~,
2. select among competing models,
3. make inferences based on one or multiple models.

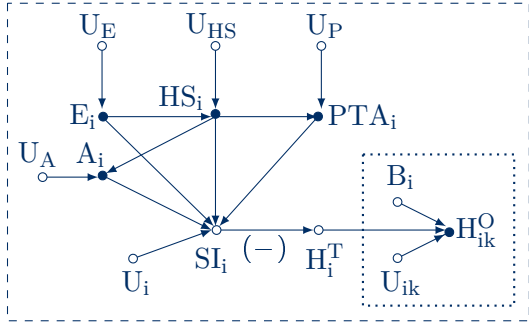
the most fitting model based on information criteria,

- WAIC [17]
- PSIS [16]



Revisiting our theory

- H_{ik} = (observed) entropy replicates
- H_i = (latent) "true" entropy
- SI_i = (latent) SI score (inversely related to H_i^T)
- A_i = "hearing" age (minimum)
- E_i = etiology of disease
- HS_i = hearing status
- PTA_i = pure tone average (standardized)
- B_i = block
- variables assumed independent, beyond the described relationships,

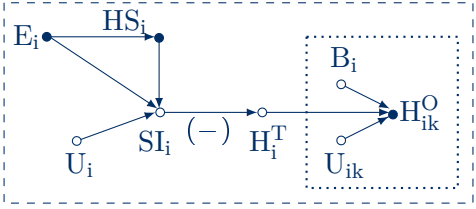


General structural diagram

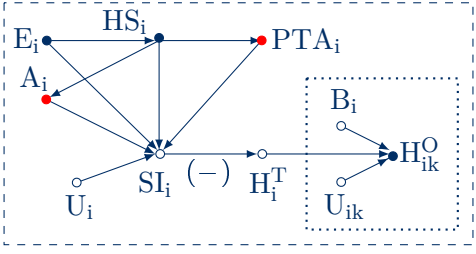
$$P(\mathbf{U}) = P(U_{ik}, U_i, U_A, U_E, U_{HS}, U_P, U_B)$$
$$= P(U_{ik})P(U_i)P(U_A)P(U_E)P(U_{HS})P(U_P)P(U_B)$$

Interested in two effects

1. total effects model inherits:
   - children's characteristics that lead to the fitting of specific apparatus,
   - the (convenience of) sample selection (fixed with post-stratification)
2. to do the last, we stratify for all variables that explain variability, ergo, use a direct effects model
3. two levels: replicates (k), children (i), denoted by discontinuous squares



(b) total effects



(a) direct effects

Competing models

```
> comp_WAIC
          WAIC    SE dWAIC   dSE pWAIC weight
E_NC2b  -620.9 42.91   0.0    NA  31.0   0.23
E_NC5b1 -619.8 42.95   1.1  0.63  31.6   0.13
E_NC5b3 -619.7 42.90   1.2  0.73  31.6   0.12
E_NC3   -619.6 42.76   1.3  2.03  34.3   0.12
E_NC6a  -619.5 42.76   1.4  2.13  34.4   0.11
E_NC5b2 -619.5 42.98   1.4  0.56  31.6   0.11
E_NC6b  -619.1 42.80   1.8  1.97  34.6   0.09
E_NC6c  -618.8 42.71   2.1  1.98  34.6   0.08
E_NC5a1 -578.6 52.54  42.3 16.95  50.4   0.00
E_NC2a  -578.6 52.65  42.4 16.99  50.5   0.00
E_NC5a3 -578.3 52.63  42.6 16.94  50.5   0.00
E_NC1   -577.4 53.06  43.5 17.56  51.5   0.00
E_NC5a2 -577.1 52.67  43.8 17.16  51.3   0.00
> comp_PSIS
          PSIS    SE dPSIS   dSE pPSIS weight
E_NC2b  -619.6 42.89   0.0    NA  31.7   0.27
E_NC5b1 -618.6 42.93   1.0  0.70  32.2   0.16
E_NC5b3 -618.5 42.88   1.1  0.77  32.1   0.16
E_NC5b2 -618.0 42.94   1.5  0.61  32.4   0.12
E_NC3   -617.5 42.79   2.1  2.04  35.4   0.09
E_NC6a  -617.1 42.78   2.5  2.14  35.6   0.08
E_NC6b  -616.5 42.78   3.1  2.05  35.9   0.06
E_NC6c  -616.5 42.68   3.1  1.96  35.8   0.06
E_NC5a1 -576.6 52.55  42.9 16.98  51.4   0.00
E_NC2a  -576.6 52.65  43.0 17.04  51.5   0.00
E_NC5a3 -575.8 52.65  43.8 17.07  51.8   0.00
E_NC1   -575.7 53.08  43.9 17.58  52.4   0.00
E_NC5a2 -575.0 52.68  44.6 17.19  52.4   0.00
```

3. References

3. References

[1] Anderson, D. [2008]. Model Based Inference in the Life Sciences: A Primer on Evidence, Springer.

[2] Betancourt, M. and Girolami, M. [2012]. Hamiltonian monte carlo for hierarchical models.
url: https://arxiv.org/abs/1312.0906v1.

[3] Boonen, N., Kloots, H., Nurzia, P. and Gillis, S. [2021]. Spontaneous speech intelligibility: early cochlear implanted children versus their normally hearing peers at seven years of age, Journal of Child Language pp. 1–26.
doi: https://doi.org/10.1017/S0305000921000714.

[4] Chamberlain, T. [1965]. The method of multiple working hypotheses, Science 148(3671): 754–759.
url: https://www.jstor.org/stable/1716334.

[5] Cohen, J. [1988]. Statistical power analysis for the behavioral sciences, Routledge.

[6] De Raeve, L. [2016]. Cochlear implants in belgium: Prevalence in paediatric and adult cochlear implantation, European Annals of Otorhinolaryngology, Head and Neck Diseases 133: S57–S60.
doi: https://doi.org/10.1016/j.anorl.2016.04.018.
url: https://www.sciencedirect.com/science/article/pii/S1879729616300813.

[7] Faes, J., De Maeyer, S. and Gillis, S. [2021]. Speech intelligibility of children with an auditory brainstem implant: a triple-case study, pp. 1–50. (submitted).

[8] Fogarty, L., Madeleine, A., Holding, T., Powell, A. and Kandler, A. [2022]. Ten simple rules for principled simulation modelling, PLOS Computational Biology 18(3): 1–8.
doi: https://doi.org/10.1371/journal.pcbi.1009917.

[9] Hernán, M. [2020]. Causal diagrams: Draw your assumptions before your conclusions.
url: https://www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your.

[10] Kruschke, J. [2014]. Doing Bayesian Data Analysis, A Tutorial with R, JAGS, and Stan, Elsevier.

[11] McElreath, R. [2020]. Statistical Rethinking: A Bayesian Course with Examples in R and STAN, Chapman and Hall/CRC.

[12] McElreath, R. [2022]. Statistical rethinking, 2022 course.
url: https://github.com/rmcelreath/stat_rethinking_2022.

[13] Pearl, J. [2019]. The seven tools of causal inference, with reflections on machine learning, Communications of the ACM 62(3): 54–60.
doi: https://doi.org/10.1177/0962280215586010.

[14] Rivera, J. [2021]. Generalized Linear Latent and Mixed Models: method, estimation procedures, advantages, and applications to educational policy., PhD thesis, KU Leuven.

[15] Sawilowsky, S. [2009]. New effect size rules of thumb, Journal of Modern Applied Statistical Methods 8(2).
doi: https://doi.org/10.22237/jmasm/1257035100.
url: http://digitalcommons.wayne.edu/jmasm/vol8/iss2/26.

[16] Vehtari, A., Simpson, D., Gelman, A., Yao, Y. and Gabry, J. [2021]. Pareto smoothed importance sampling.
url: https://arxiv.org/abs/1507.02646.

[17] Watanabe, S. [2013]. A widely applicable bayesian information criterion, Journal of Machine Learning Research 14 14: 867–897.
url: https://dl.acm.org/doi/10.5555/2567709.2502609.

[18] Winkler, R. [1967]. The assessment of prior distributions in bayesian analysis, Journal of the American Statistical Association 62(319): 776–800.
doi: https://doi.org/10.1080/01621459.1967.10500894.
url: https://www.tandfonline.com/doi/abs/10.1080/01621459.1967.10500894.