



University of Antwerp
| Faculty of Social Sciences

Speech intelligibility measurement

A latent variable approach on transcription of utterances

Jose Rivera

May 30, 2022

What are we going to talk about?

1 Preliminars

- Research question
- Research hypothesis production

2 Research hypothesis procedure

- Estimand and process model
- Synthetic data generation
- Statistical model design and testing
- Apply statistical model to data

3 References

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 we 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 $H \rightarrow 0$,
- utterances with low agreement are less intelligible, and therefore $H \rightarrow 1$.

1. Preliminars

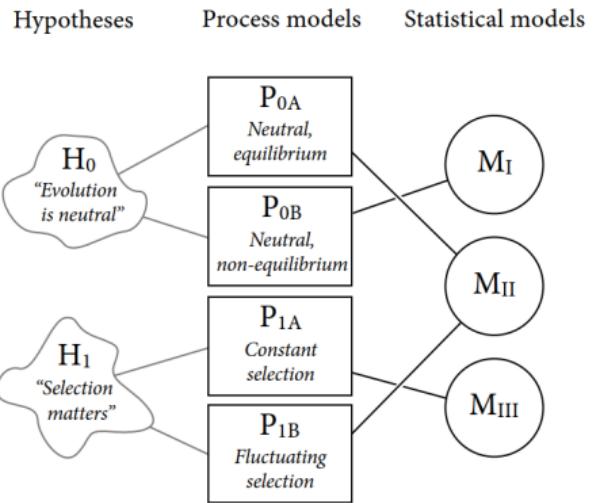
Research hypothesis production

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

^aFigure extracted from McElreath [11]



Research hypothesis production¹

Well known challenges

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

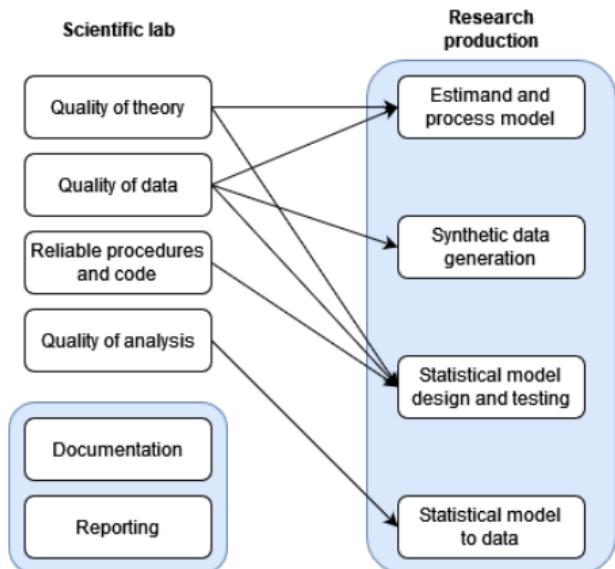
Known challenges in our research;

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

¹Hernán [9], lesson 4

Research hypothesis schematics²

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



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

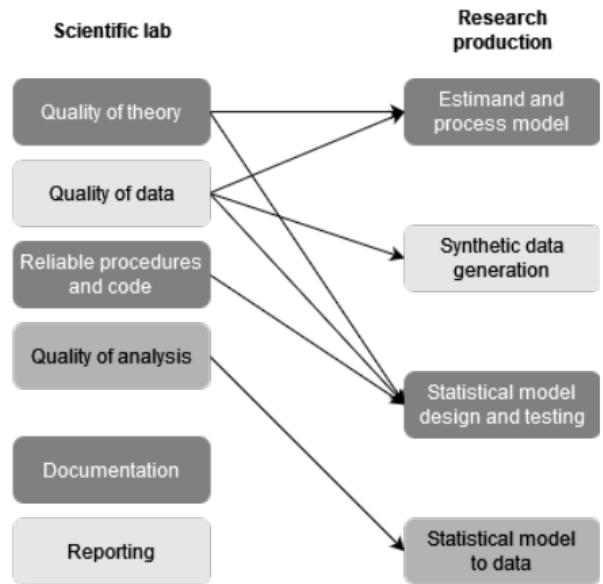
Why do we need to follow this?

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.



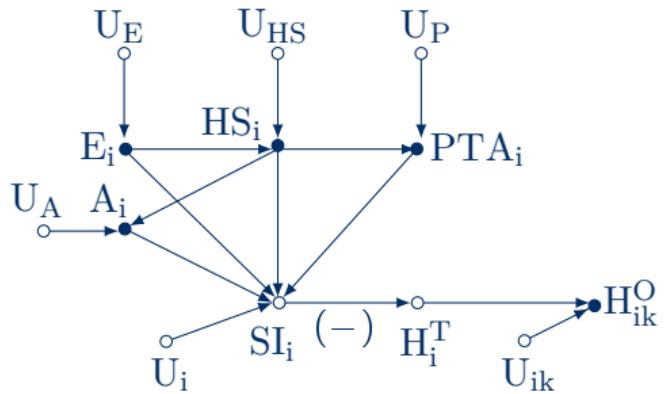
2. Research hypothesis procedure

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,

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



General structural diagram

Probabilistic (causal) model

First form

$$H_{ik}^O \leftarrow f(H_i^T, U_{ik})$$

$$H_i^T \leftarrow f(SI_i)$$

$$SI_i \leftarrow f(HS_i, E_i, A_i, PTA_i, U_i)$$

$$HS_i \leftarrow f(U_{HS})$$

$$A_i \leftarrow f(U_A)$$

$$E_i \leftarrow f(U_E)$$

$$PTA_i \leftarrow f(U_P)$$

$$U \sim P(U)$$

(a) general structural model

$$H_{ik}^O \sim \text{BetaProp}(H_i^T, M_{ik})$$

$$H_i^T = \text{inv_logit}(-SI_i)$$

$$SI_i \sim \text{Normal}(\mu_{SI}, \sigma_{Ui} +)$$

$$\begin{aligned}\mu_{SI} = & \alpha + \alpha_{E[i], HS[i]} + \beta_P PTA_i \\ & + \beta_{A, HS[i]} (A_i - \bar{A})\end{aligned}$$

$$HS_i \sim \text{data}$$

$$A_i \sim \text{data}$$

$$E_i \sim \text{data}$$

$$PTA_i \sim \text{data}$$

$$U \sim \text{unobservable}$$

(a) general probabilistic model

Probabilistic (causal) model

First form

Notice,

- $\alpha, \alpha_{HS[i]}, \alpha_{E[i]}, \beta_{A,HS[i]}, \beta_P$ 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}

$$\begin{aligned} H_{ik}^O &\sim \text{BetaProp}(H_i^T, M_{ik}) \\ H_i^T &= \text{inv_logit}(-SI_i) \\ SI_i &\sim \text{Normal}(\mu_{SI}, \sigma_{Ui}) \\ \mu_{SI} &= \alpha + \alpha_{E[i], HS[i]} + \beta_P PTA_i \\ &\quad + \beta_{A, HS[i]}(A_i - \bar{A}) \\ HS_i &\sim \text{data} \\ A_i &\sim \text{data} \\ E_i &\sim \text{data} \\ PTA_i &\sim \text{data} \\ U &\sim \text{unobservable} \\ \end{aligned}$$

(a) general probabilistic model

Probabilistic (causal) model

Express variability in BetaProp

$$H_{ik}^O \sim \text{BetaProp}(H_i^T, M_{ik})$$

$$H_i^T = \alpha / (\alpha + \beta)$$

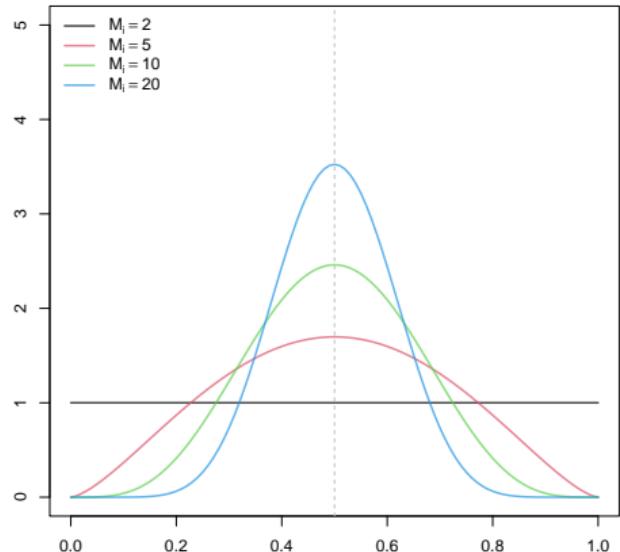
$$M_{ik} = \alpha + \beta$$

$$\alpha = H_i^T \cdot M_{ik}$$

$$\beta = (1 - H_i^T) \cdot M_{ik}$$

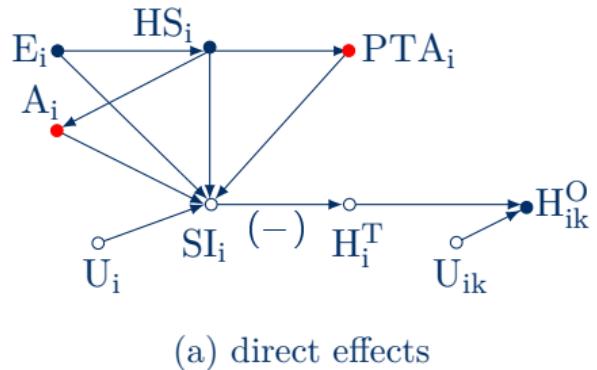
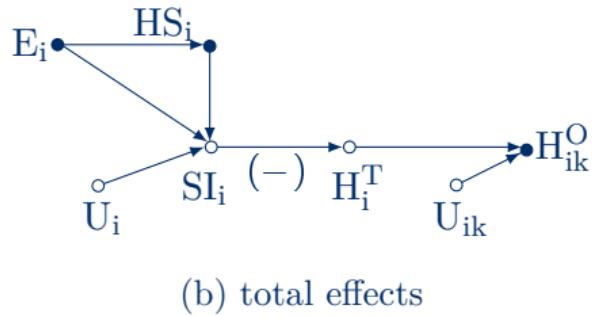
$$\alpha = 0.5 \cdot 2$$

$$\beta = (1 - 0.5) \cdot 2$$



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



Probabilistic (causal) model

Second form

$$H_{ik}^O \leftarrow f(H_i^T, U_{ik})$$

$$H_i^T \leftarrow f(SI_i)$$

$$SI_i \leftarrow f(HS_i, E_i, A_i, PTA_i, U_i)$$

$$HS_i \leftarrow f(U_{HS})$$

$$A_i \leftarrow f(U_A)$$

$$E_i \leftarrow f(U_E)$$

$$PTA_i \leftarrow f(U_P)$$

$$U \sim P(U)$$

(a) general structural model

$$H_{ik}^O \sim BetaProp(H_i^T, M_{ik})$$

$$H_i^T = inv_logit(-SI_i)$$

$$\begin{aligned} SI_i = & a_i + \alpha + \alpha_{E[i], HS[i]} + \beta_P PTA_i \\ & + \beta_{A, HS[i]} (A_i - \bar{A}) \end{aligned}$$

$$a_i \sim Normal(0, \sigma_{Ui})$$

$$HS_i \sim data$$

$$A_i \sim data$$

$$E_i \sim data$$

$$PTA_i \sim data$$

$$U \sim unobservable$$

(a) general probabilistic model

2. Research hypothesis procedure

Synthetic data generation

Idealized data³

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 include
sim_save=NULL, # file_save need to include
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 ch
          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,-0.1,0.05),
                           c(0, seq(0.1,-0.1,0.05),
                           ncol=3, byrow=F) )
```

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

Idealized data

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

1. E might have a full interaction with HS
(it can be easily implemented)
2. $a_{HS}[j] - a_{HS}[i] \approx -0.4$,
NH vs HI/CI
(might depend on E)
3. $bP = -0.1$, per PTA unit
 $(+10 \text{ PTA units} \Rightarrow -1 \text{ logit})$,
4. $bA \approx 0.15$, per A unit $> \min(A)$
 $(+10 \text{ A units} \Rightarrow +1.5 \text{ logits})$
(might depend on HS)

```
sim_name=NULL, # file_name need to include
sim_save=NULL, # file_save need to include
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 ch
          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,-0.1, -0.1),
                            c(0, seq(0.1,-0.1, -0.1),
                            ncol=3, byrow=F) )
```

Idealized data

- variables are generated in a random fashion
- between children SI variability are defined by the random effects

```
# 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 (H_t) 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: negative)

# 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$s_-
}
dT$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)

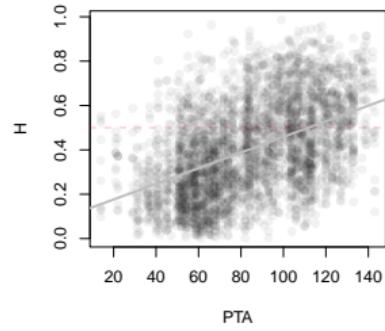
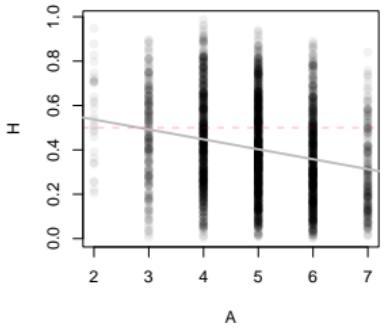
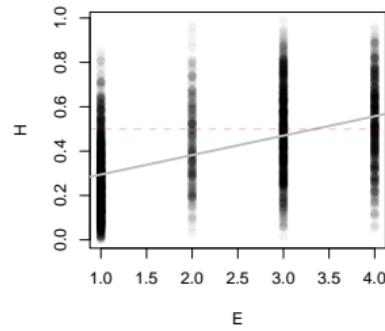
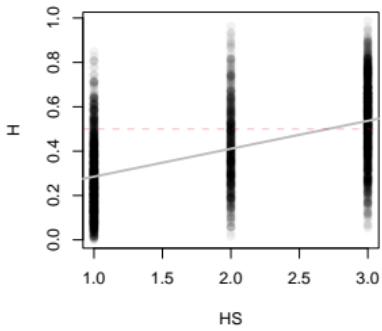
# generating observed H
# i=1
if(!is.null(seed)){
  set.seed(seed-2)
}
for(i in 1:I){

  # identify data
  idx = do$child_id == i

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

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

Example



2. Research hypothesis procedure

Statistical model design and testing

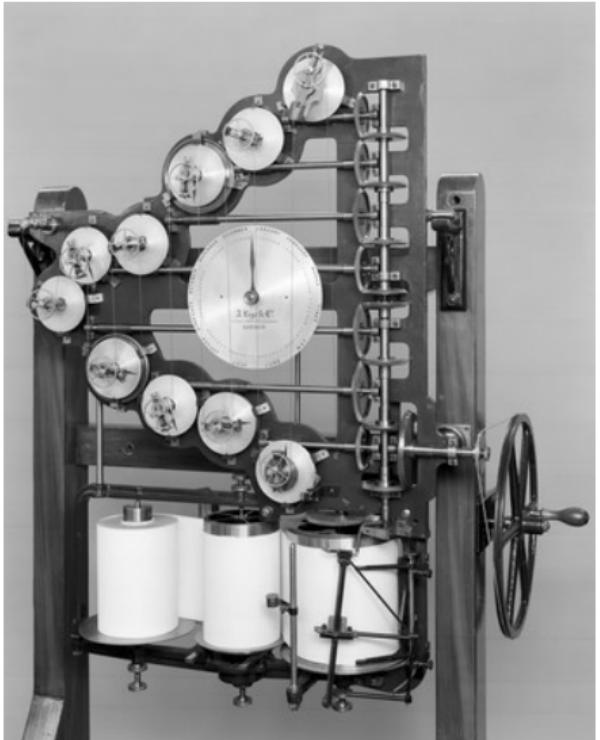
Model design and test⁴

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

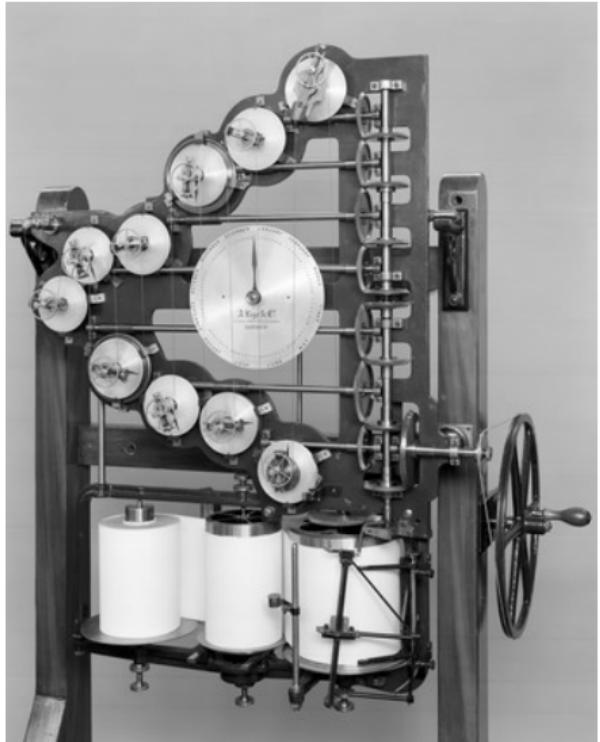


⁴Following Fogarty et al. [8]

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 child)

  SI = a + re_i;          // linear predictor
  Ht = inv_logit(-SI);   // average entropy (SI - 1)
}
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 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 -> Ht)
}

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 -> Ht)  
}  
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

$$a_i \sim \text{Normal}(\mu_a, \sigma_a)$$

$$M_i \sim \text{LogNormal}(\mu_M, \sigma_M)$$

$$\alpha \sim \text{Normal}(0, 0.2)$$

$$\alpha_{HS[i]} \sim \text{Normal}(0, 0.3)$$

$$\alpha_{E[i]} \sim \text{Normal}(0, 0.3)$$

$$\beta_{A,HS[i]} \sim \text{Normal}(0, 0.3)$$

$$\beta_P \sim \text{Normal}(0, 0.3)$$

Hyper-priors

$$\mu_a \sim \text{Normal}(0, 0.2)$$

$$\sigma_a \sim \text{Exp}(1)$$

$$\mu_M \sim \text{Normal}(0, 0.5)$$

$$\sigma_M \sim \text{Exp}(1)$$

Prior predictive

Undesired assumptions can easily creep in **non-linear models**^a

Example:

(black line)

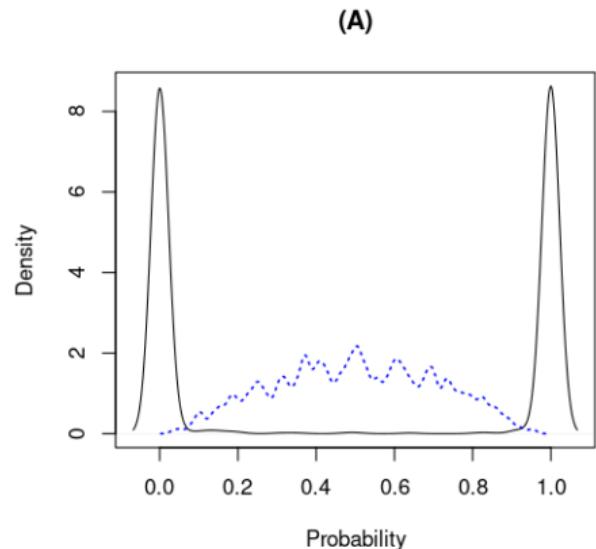
$$\theta \sim N(0, 100)$$

$$\text{logit}(p) = \theta$$

(blue line)

$$\theta \sim N(0, 1)$$

$$\text{logit}(p) = \theta$$



^aFigure extracted from Rivera [14].

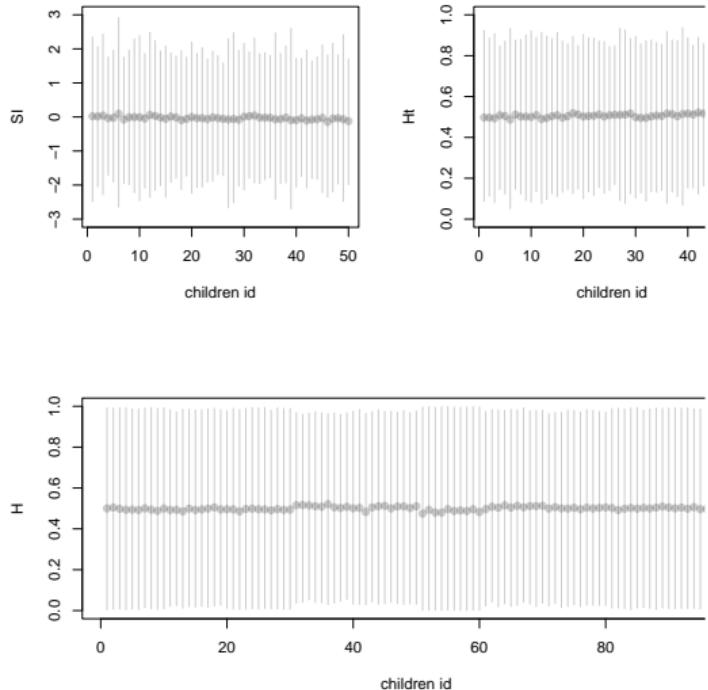
Prior predictive

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

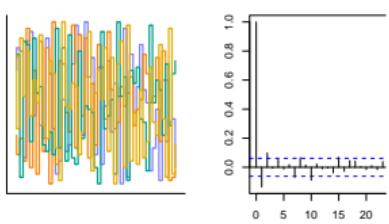
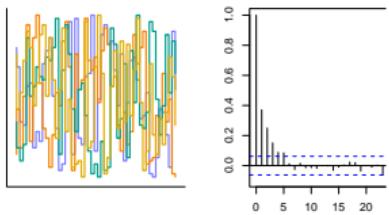
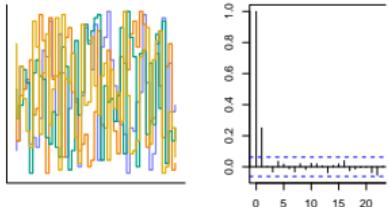
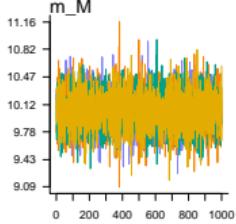
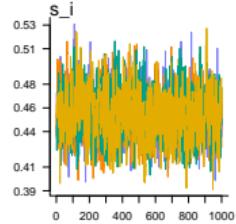
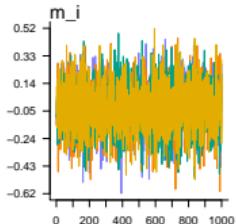


“Health” of MCMC chains

The MCMC chains achieve,

- good convergence
- good mixing
- lack of autocorrelation

same results on the `n_eff` and `RHat` statistics.

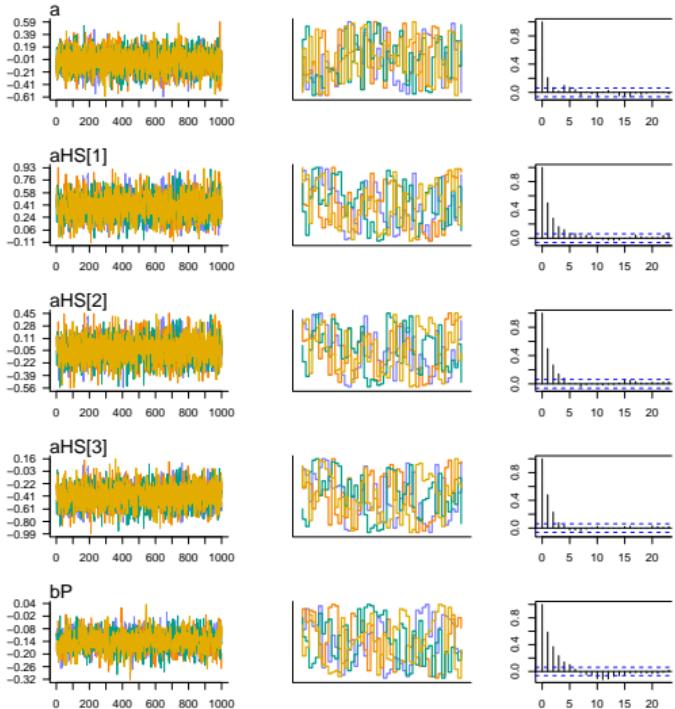


“Health” of MCMC chains

The MCMC chains achieve,

- good convergence
- good mixing
- lack of autocorrelation

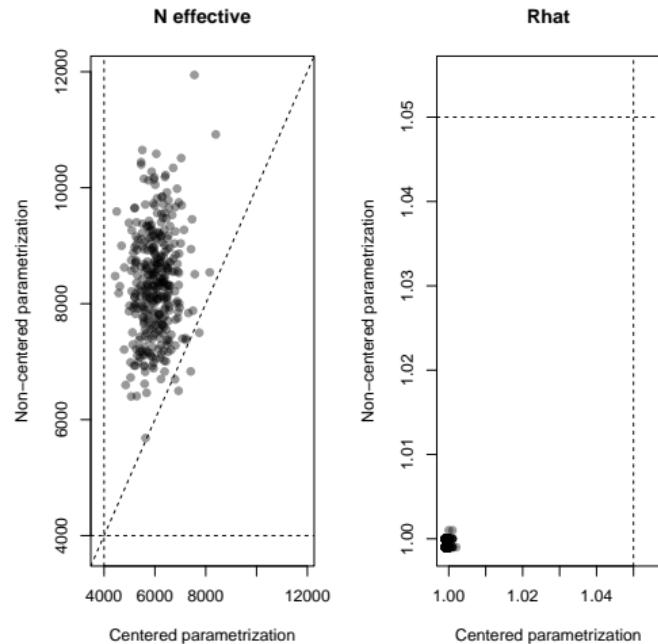
same results on the `n_eff` and `RHat` statistics.



“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)



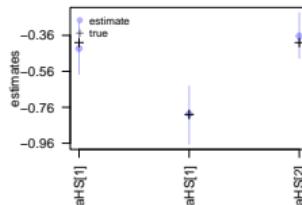
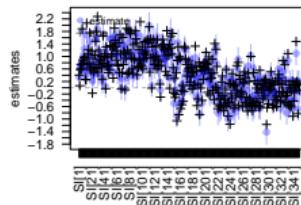
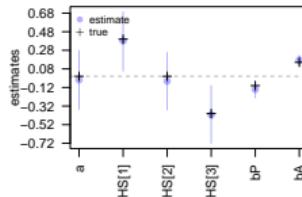
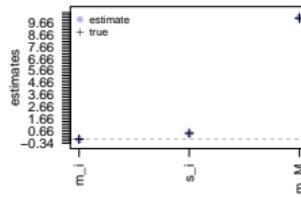
Parameter recovery

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

^arandom effects causal model, with $M = 10$



Posterior predictive

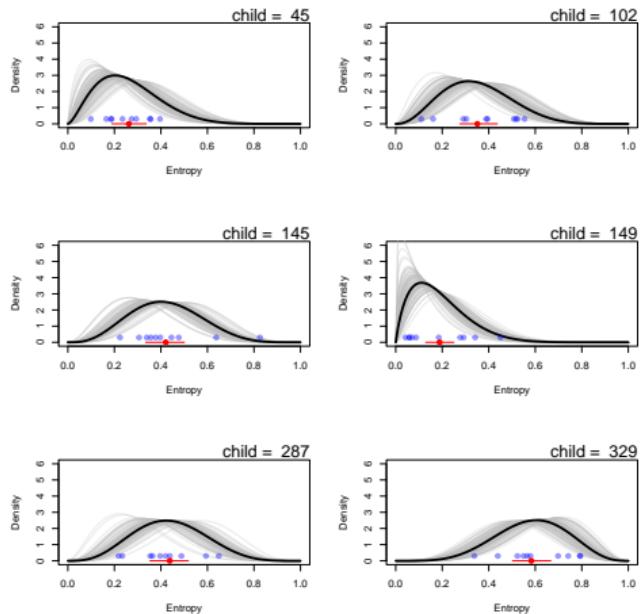
But how well reproduces the data?,
with less working assumptions^a

- 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])

^arandom effects causal model, with M = 10



Power

Equivalent prior sampling method [18, 10]

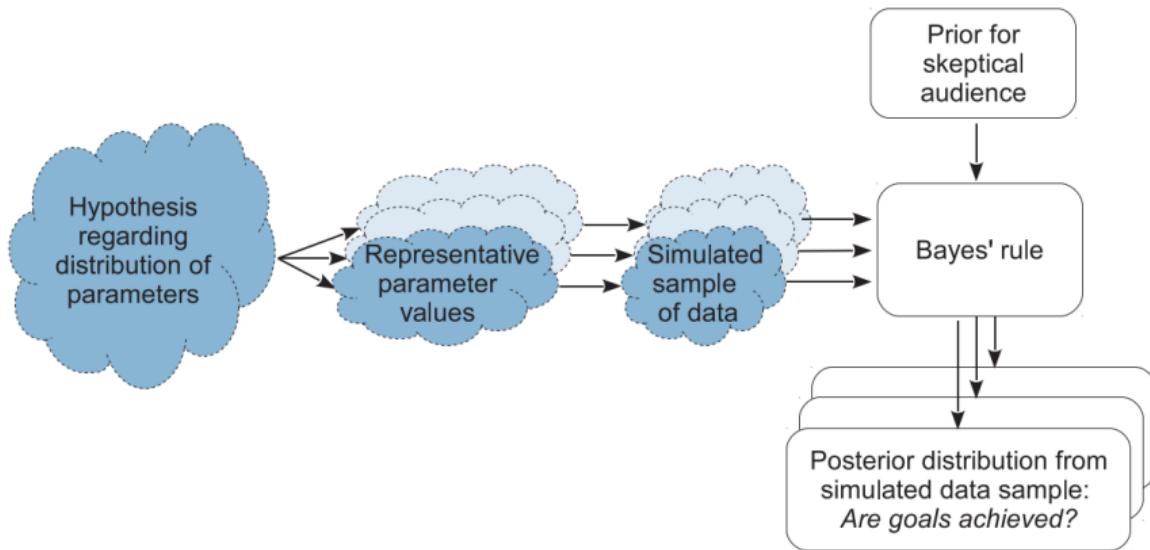
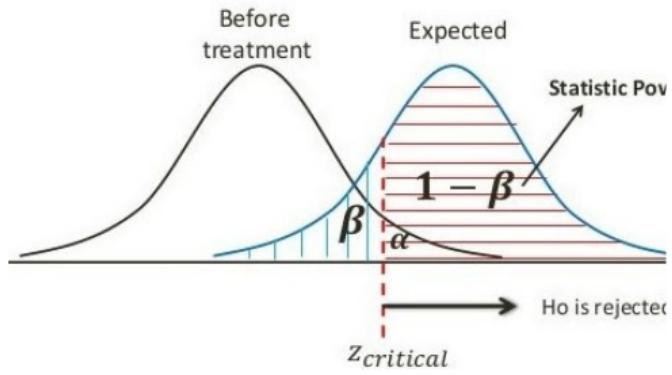


Figure: Flow of information in a power analysis, in which simulated data come from random hypothetical parameters. Extracted from Kruschke [10].

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)

```
(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

^ause ROPE [10], and effect sizes [5, 15]

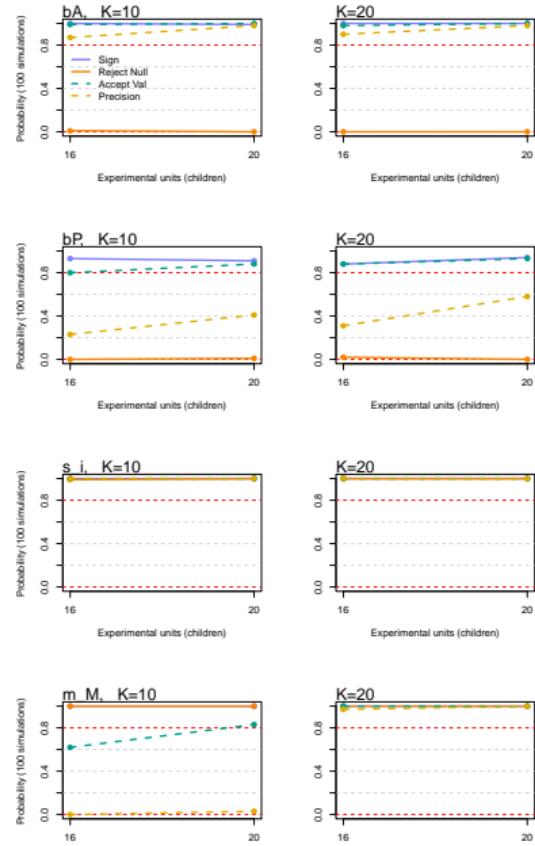
Power results

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)



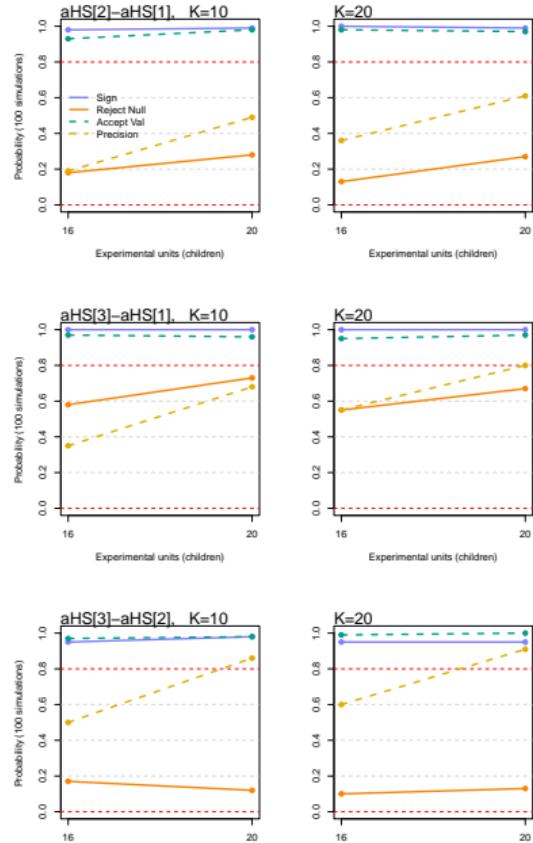
Power results

For group contrasts,
assuming medium differences (≈ 0.4)

- reject the null hypothesis,
 $\approx 80\%$ 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 ($K = 20$)
- group contrasts are easier to identify (larger sample size per comparison)



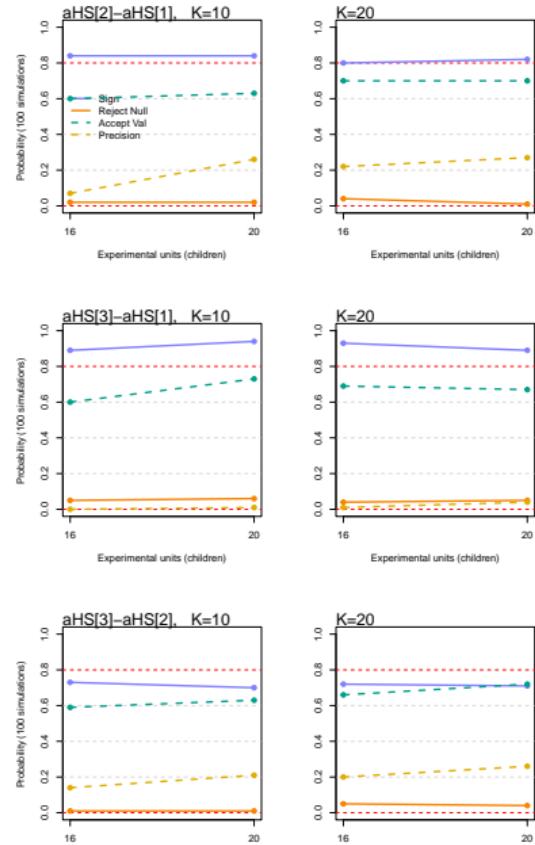
Power results

For group contrasts,
assuming small differences (≈ 0.15)

- reject the null hypothesis, none,
- affirm predicted value, $\approx 60\%$ for all contrast,
- achieve precision in estimate, too low for this data size

Notice,

- not much difference with 2x comparisons ($K = 20$)
- group contrasts are easier to identify (larger sample size per comparison)



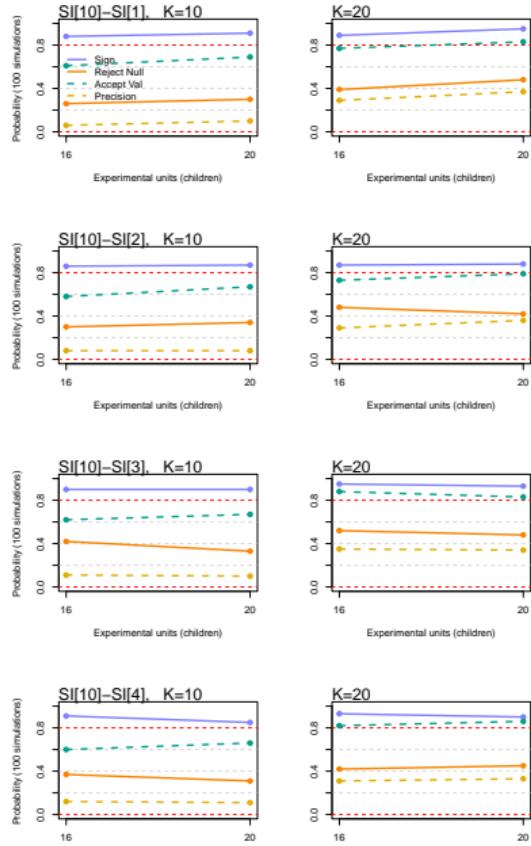
Power results

For individual contrasts,
(either with ≈ 0.4 or ≈ 0.15 differences)

- 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)

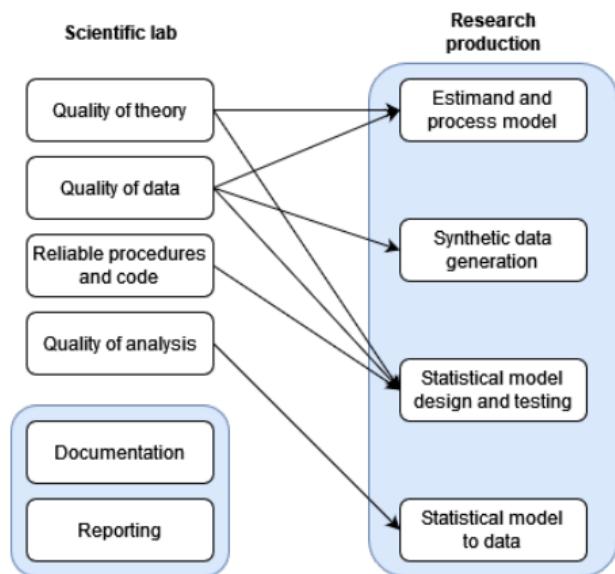


2. Research hypothesis procedure

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



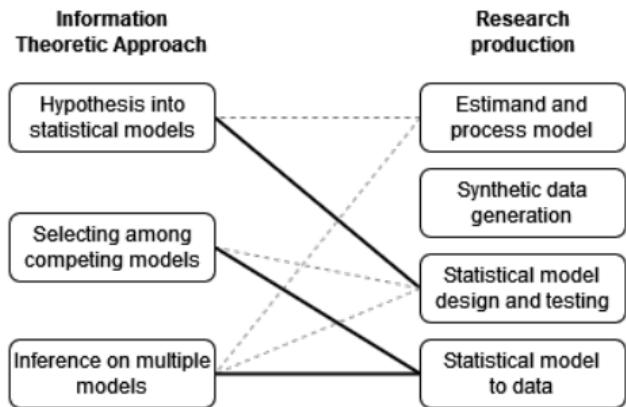
Information Theoretic Approach [1, 4]

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,

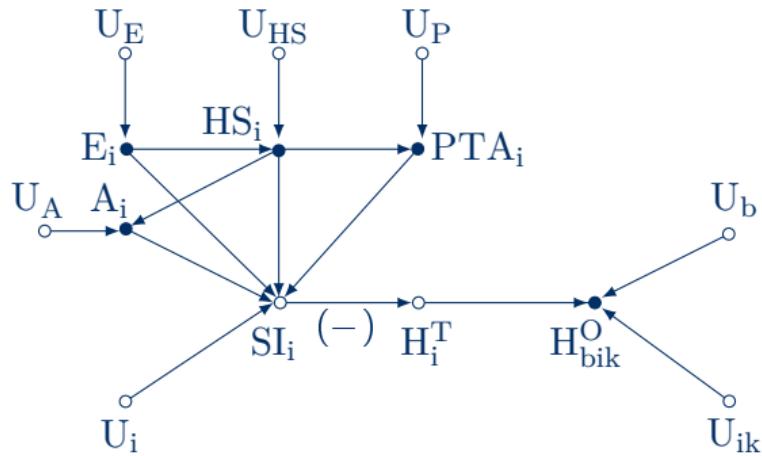
- WAIC [17]
- PSIS [16]



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)
- variables assumed independent,
beyond the described relationships,

$$\begin{aligned} P(\mathbf{U}) &= P(U_{ik}, U_i, U_A, U_E, U_{HS}, U_P, B_{bk}) \\ &= P(U_{ik})P(U_i)P(U_A)P(U_E)P(U_{HS})P(U_P)P(B_{bk}) \end{aligned}$$

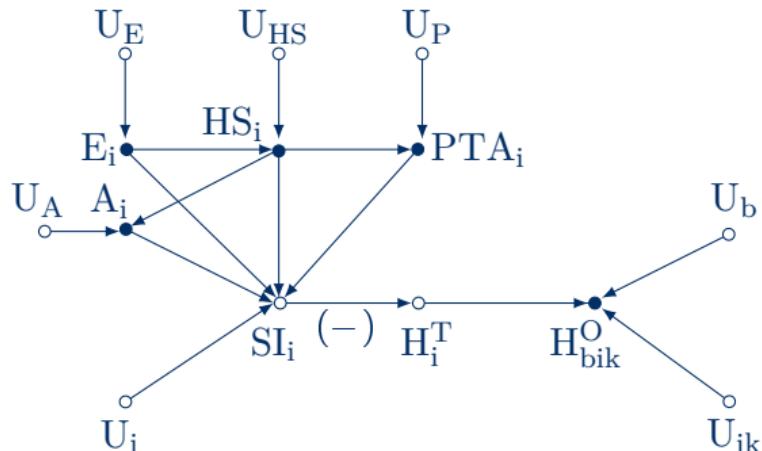


General structural diagram

Levels of variability

Three levels ;

- U_i = children's variability
($i = 1, \dots, I$, and $I = 32$)
- U_k = replicates' variability
($k = 1, \dots, K$, and $K = 10$)
- U_b = block's variability
($b = 1, \dots, B$, and $B = 5$)



General structural diagram

Competing models

We can notice,

- No evidence in favor of models with $M = 10$, i.e. weight = 0
models: E_NC1, E_NC2a,
E_NC5a1, E_NC5a2, E_NC5a3.
- Small evidence in favor of “robust” models, i.e. one M per child
models: E_NC6a, E_NC6b,
E_NC6c.
- Small evidence in favor of “robust” model without interaction
models: E_NC3.
(it might assess as ME the lack of explanation at the child level)

	WAIC	SE	dWAIC	dSE	pWAIC	weight
E_NC2b	-621.0	42.99	0.0	NA	31.1	0.18
E_NC5b3	-621.0	42.98	0.1	0.78	31.2	0.18
E_NC5b2	-620.7	43.07	0.4	0.48	31.2	0.15
E_NC5b1	-620.2	42.91	0.8	0.76	31.2	0.12
E_NC6a	-620.1	42.70	0.9	2.21	34.2	0.12
E_NC6b	-619.9	42.77	1.1	2.06	34.4	0.10
E_NC3	-619.2	42.81	1.8	2.18	34.8	0.07
E_NC6c	-618.8	42.65	2.2	2.17	35.0	0.06
E_NC5a1	-578.2	52.51	42.8	16.85	50.7	0.00
E_NC5a2	-578.2	52.63	42.8	16.95	50.7	0.00
E_NC2a	-577.6	52.59	43.5	16.97	51.0	0.00
E_NC1	-576.8	52.98	44.3	17.40	51.8	0.00
E_NC5a3	-575.8	52.58	45.3	16.96	52.1	0.00
	PSIS	SE	dPSIS	dSE	pPSIS	weight
E_NC2b	-619.8	42.98	0.0	NA	31.8	0.22
E_NC5b3	-619.7	42.97	0.0	0.82	31.8	0.22
E_NC5b2	-619.2	43.05	0.5	0.51	32.0	0.17
E_NC5b1	-619.0	42.89	0.7	0.81	31.8	0.15
E_NC6a	-617.8	42.70	2.0	2.19	35.4	0.08
E_NC6b	-617.7	42.77	2.1	2.07	35.5	0.08
E_NC3	-616.5	42.81	3.3	2.19	36.1	0.04
E_NC6c	-616.3	42.64	3.4	2.13	36.2	0.04
E_NC5a2	-576.0	52.63	43.7	17.00	51.7	0.00
E_NC5a1	-575.9	52.54	43.9	16.94	51.9	0.00
E_NC2a	-575.3	52.59	44.4	17.03	52.1	0.00
E_NC1	-574.6	52.97	45.1	17.44	52.9	0.00
E_NC5a3	-573.8	52.61	46.0	17.01	53.1	0.00

Competing models

further,

- Highest evidence in favor of “simplest” model, i.e. one global M and no interaction, models: E_NC2b.
- However, is indistinguishable from the interaction models especially E_NC5b3 models: E_NC5b1, E_NC5b2.

	WAIC	SE	dWAIC	dSE	pWAIC	weight
E_NC2b	-621.0	42.99	0.0	NA	31.1	0.18
E_NC5b3	-621.0	42.98	0.1	0.78	31.2	0.18
E_NC5b2	-620.7	43.07	0.4	0.48	31.2	0.15
E_NC5b1	-620.2	42.91	0.8	0.76	31.2	0.12
E_NC6a	-620.1	42.70	0.9	2.21	34.2	0.12
E_NC6b	-619.9	42.77	1.1	2.06	34.4	0.10
E_NC3	-619.2	42.81	1.8	2.18	34.8	0.07
E_NC6c	-618.8	42.65	2.2	2.17	35.0	0.06
E_NC5a1	-578.2	52.51	42.8	16.85	50.7	0.00
E_NC5a2	-578.2	52.63	42.8	16.95	50.7	0.00
E_NC2a	-577.6	52.59	43.5	16.97	51.0	0.00
E_NC1	-576.8	52.98	44.3	17.40	51.8	0.00
E_NC5a3	-575.8	52.58	45.3	16.96	52.1	0.00
	PSIS	SE	dPSIS	dSE	pPSIS	weight
E_NC2b	-619.8	42.98	0.0	NA	31.8	0.22
E_NC5b3	-619.7	42.97	0.0	0.82	31.8	0.22
E_NC5b2	-619.2	43.05	0.5	0.51	32.0	0.17
E_NC5b1	-619.0	42.89	0.7	0.81	31.8	0.15
E_NC6a	-617.8	42.70	2.0	2.19	35.4	0.08
E_NC6b	-617.7	42.77	2.1	2.07	35.5	0.08
E_NC3	-616.5	42.81	3.3	2.19	36.1	0.04
E_NC6c	-616.3	42.64	3.4	2.13	36.2	0.04
E_NC5a2	-576.0	52.63	43.7	17.00	51.7	0.00
E_NC5a1	-575.9	52.54	43.9	16.94	51.9	0.00
E_NC2a	-575.3	52.59	44.4	17.03	52.1	0.00
E_NC1	-574.6	52.97	45.1	17.44	52.9	0.00
E_NC5a3	-573.8	52.61	46.0	17.01	53.1	0.00

Selecting models

“simplest” model (E_NC2b) provides (preliminar) evidence on,

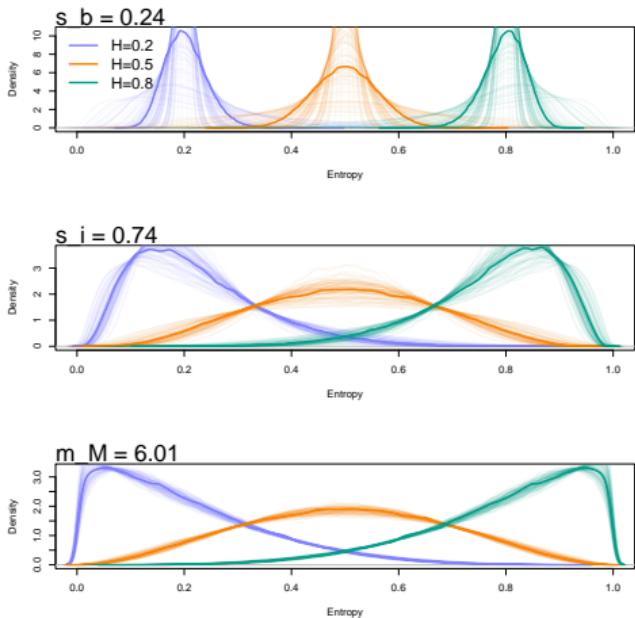
- the higher the unaided PTA the lower the child’s SI (bP[2])
(based on power analysis, we can be sure is a small effect)
- no apparent statistical difference between NH and HI/CI children
(but this requires a CONTRAST)
- for each “hearing” year, the SI increases in ≈ 0.40 logits
(effect larger than the assumed in power analysis)

	mean	sd	CI_lower	CI_upper
a	0.179	0.189	-0.197	0.554
bP[1]	0.000	0.308	-0.599	0.601
bP[2]	-0.117	0.166	-0.439	0.207
aHS[1]	0.284	0.235	-0.178	0.761
aHS[2]	0.116	0.217	-0.303	0.537
bA	0.432	0.141	0.154	0.715
m_i	0.174	0.183	-0.184	0.539
s_i	0.743	0.117	0.549	1.014
m_b	-0.180	0.183	-0.539	0.174
s_b	0.243	0.195	0.026	0.785
m_M	6.013	0.510	5.075	7.041

Selecting models

additionally the “simplest” model tells us that,

- it is hard to determine how the variability is “explained” just by looking at the parameters s_b , s_i , m_M
- transforming the parameters into $[0, 1]$ entropy scale:
 - block level (s_b) has the lowest variability,
 - the individual level (s_i) follows,
 - the measurement level variability (m_M) has the highest



Selecting models

however, the “interaction” model (E_NC5b3) shows similar results on,

- the small (still non-significant) effects of the unaided PTA on the child’s SI
- similar explained variability across levels and blocks
(similar to the “simplest” model)
- but “mild” evidence of prevalent interactions,
 - SI means for HI/CI per E, aEHS[2, 2] (Genetic) vs aEHS[3, 2] (CMV)
 - different SI evolution for NH vs HI/CI children, per unit A (bAHS),

	mean	sd	CI_lower	CI_upper
a	0.217	0.179	-0.142	0.562
bP[1]	-0.001	0.306	-0.615	0.588
bP[2]	-0.122	0.173	-0.457	0.220
aEHS[1,1]	0.183	0.278	-0.358	0.726
aEHS[1,2]	-0.002	0.294	-0.594	0.572
aEHS[2,1]	0.000	0.298	-0.591	0.600
aEHS[2,2]	0.212	0.241	-0.259	0.684
aEHS[3,1]	0.002	0.297	-0.577	0.572
aEHS[3,2]	0.077	0.245	-0.402	0.547
aEHS[4,1]	0.001	0.304	-0.583	0.609
aEHS[4,2]	0.007	0.269	-0.522	0.530
bAHS[1]	0.435	0.157	0.127	0.745
bAHS[2]	0.237	0.177	-0.119	0.596
m_i	0.216	0.190	-0.161	0.584
s_i	0.737	0.118	0.541	1.002
m_b	-0.215	0.182	-0.576	0.150
s_b	0.259	0.228	0.032	0.847
m_M	5.997	0.515	5.006	7.031

Selecting models

within the “interaction” model,

- the size of the data within groups from combinations of E and HS, does not allow to reject the contrasts’ null hypothesis,
- similar result is observed on the bAHS contrast (because the effect is small, compared to children’s variability)
- but we still observe differences between NH and HI/CI, and even within HI/CI by E,
- therefore we decide to keep the (E_NC5b3) model

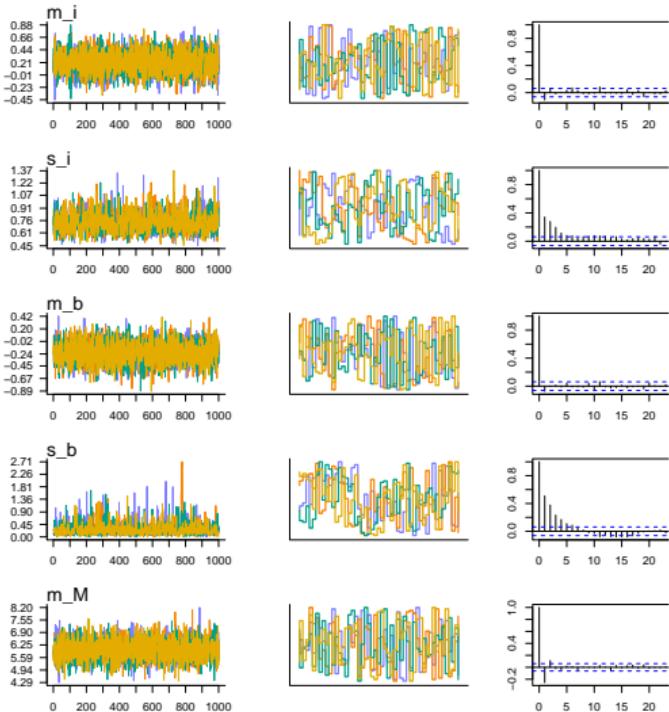
		mean	sd	CI_lower	CI_upper
aEHS[2,2]	- aEHS[1,1]	0.029	0.352	-0.673	0.723
aEHS[3,2]	- aEHS[1,1]	-0.106	0.365	-0.823	0.607
aEHS[4,2]	- aEHS[1,1]	-0.176	0.380	-0.906	0.568
aEHS[3,2]	- aEHS[2,2]	-0.135	0.299	-0.719	0.453
aEHS[4,2]	- aEHS[2,2]	-0.205	0.344	-0.888	0.450
aEHS[4,2]	- aEHS[3,2]	-0.070	0.344	-0.744	0.612
bAHS[2]	- bAHS[1]	-0.197	0.208	-0.600	0.213

“Health” of MCMC chains

The MCMC chains achieve,

- good convergence
- good mixing
- lack of autocorrelation

same results on the `n_eff` and `RHat` statistics.

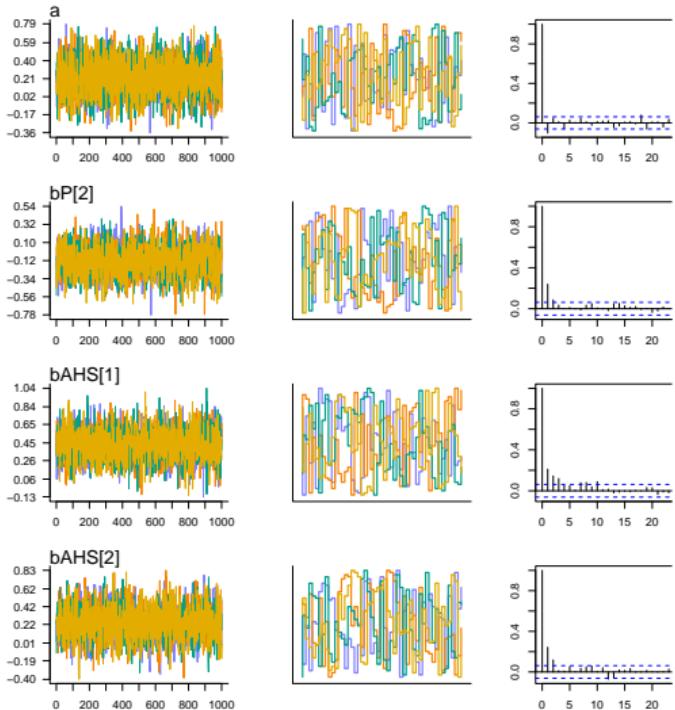


“Health” of MCMC chains

The MCMC chains achieve,

- good convergence
- good mixing
- lack of autocorrelation

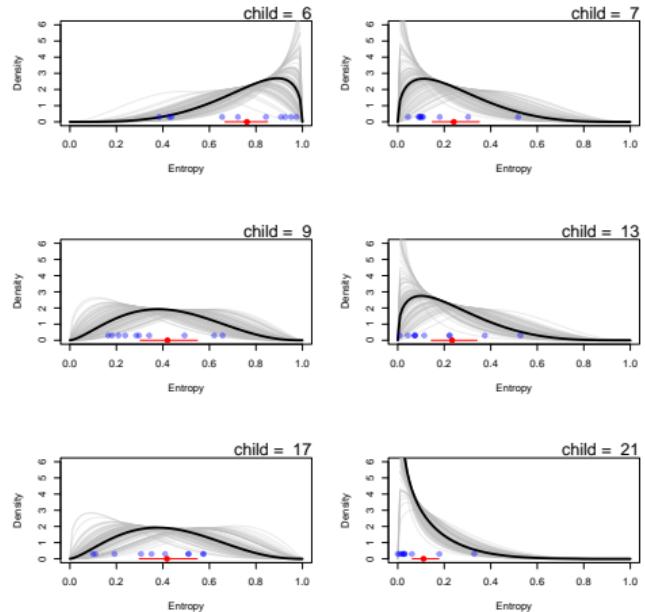
same results on the `n_eff` and `RHat` statistics.



Posterior predictive

But how well reproduces the data?,

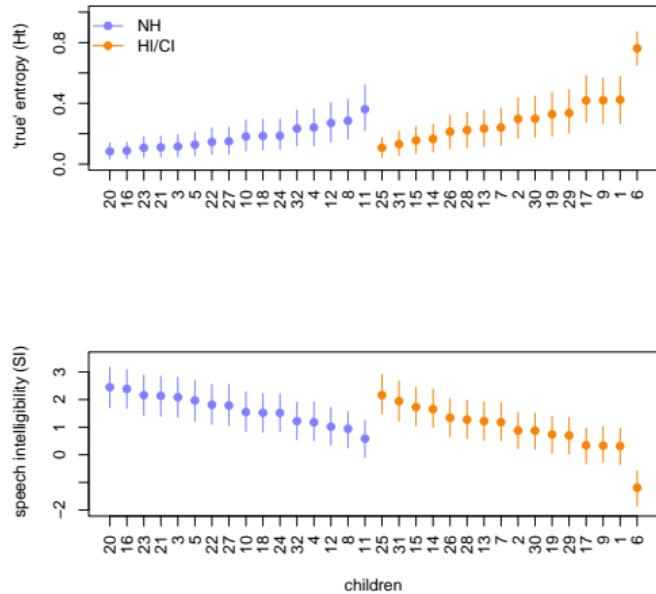
- captures the variability of the replicates,



Posterior predictive

and besides,

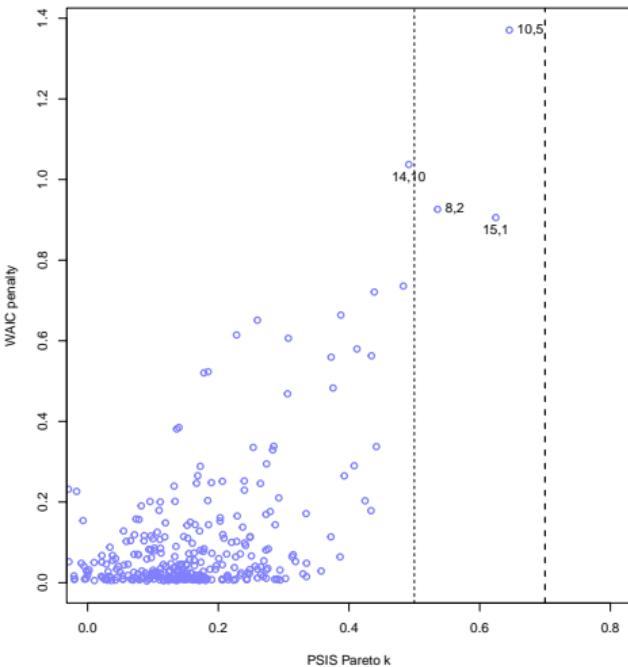
- provides a “true” H (entropy) and SI measures
 - we can also make contrast of SI measures per child



High leverage observations

“interaction” model has three (3) “high” leverage observations,
(observations that “drag” the estimation)

- one has a high “drag” ($k > 0.7$),
- the pair (child, utterance) reveals,
 - (almost) all cases cases have $H_{ik} = 0$
(next minimum value is ≈ 0.01)
 - child 15 is “young” HI/CI with “perfect” H_{bik} ,
(the utterance is a well developed word, for him/her?)
 - no other apparent relationship among them



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>.