

lecture 1. Bayesian regression

Ülo Maiväli

uncertainty is measured by posterior probability distribution



a simple generative statistical model

- $y \sim N(\mu, \sigma)$ or
 - $y \sim f(\theta)$
- y – predicted variable
- N – normal distribution
- μ, σ – parameters
 - θ – vector of parameters
 - f – a function
 - tilde means that data are drawn stochastically from a normal distribution, whose parameters are fixed (fitted on data)

The basic story of Bayesian statistics:
two probability distributions are converted into
a single narrower distribution – the posterior

1. Initial beliefs concerning a parameter of interest (based on objective evidence, or subjective judgment, or a combination thereof) are converted into a prior distribution.
2. data is modelled as a likelihood function for the parameter
3. the normalized product of the prior and the likelihood form the posterior distribution

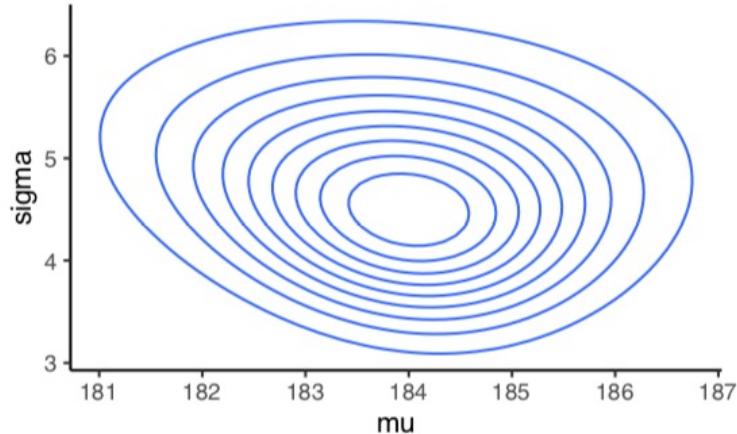
Bayesian inference uses the information provided by data about parameter(s), the *likelihood*, to update a *prior* state of beliefs about parameter(s) to become a *posterior* state of beliefs about parameter(s).

$$p(\mu | D) \propto p(D | \mu) \cdot p(\mu)$$

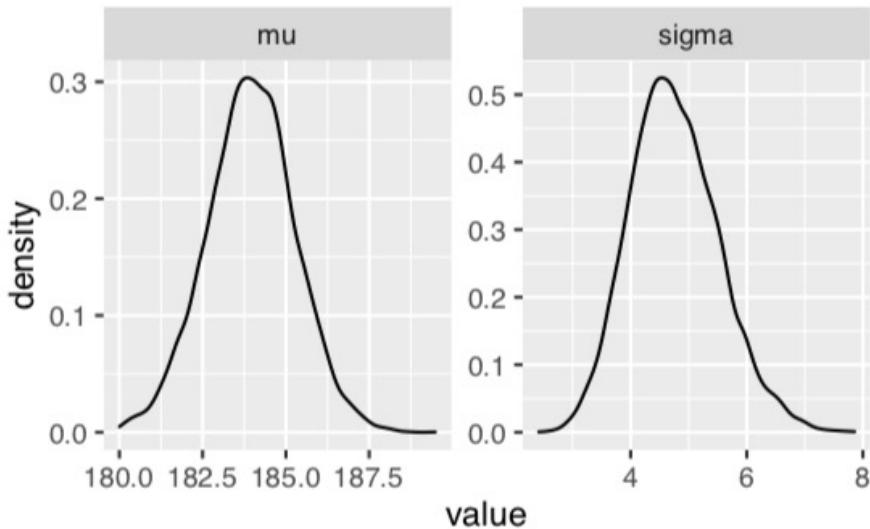
- μ - parameter(s), D - data,
- $p(\mu | D)$ - the posterior or the probability of μ given the data,
- $p(D | \mu)$ - the likelihood or the probability of the data given μ ,
- $p(\mu)$ - the prior or the apriori probability of μ .
- the data are used to update the prior belief is by examining the likelihood of the data given certain value(s) of the parameter(s).
- Ideally, one would like to assess this likelihood for every single combination of parameter values. MCMC draws a sequence of samples from the posterior.

- Lets generate some heights of US presidential candidates. assume that human growth is a deterministic process that depends on a complex mix of many factors.
- we do not understand this process well enough to directly use it to generate prospective heights of future presidents. We think that human height is approximately normally distributed (conditional on sex, age, etc.), so we could model presidential heights as a normal distribution where each possible height is given a probability of occurring.
- we could generate new heights from this model proportionally to these probabilities. But first we have to fix the shape of the normal distribution so that we know the relevant probabilities. we fix the values of 2 parameters of the normal distribution, μ and σ . Luckily we have data on the heights of the last 12 presidents:
 - heights <- **tibble**(value = c(183, 192, 182, 183, 177, 185, 188, 188, 182, 185, 188, 182))
- As μ is identical to the mean of the normal and σ to standard deviation (SD), we can simply calculate those and then use these numbers to generate new heights.
 - Data mean is: (Mean <- **mean**(heights\$value)) ## [1] 184.5833
 - SD is: (SD <- **sd**(heights\$value)) ## [1] 3.964807

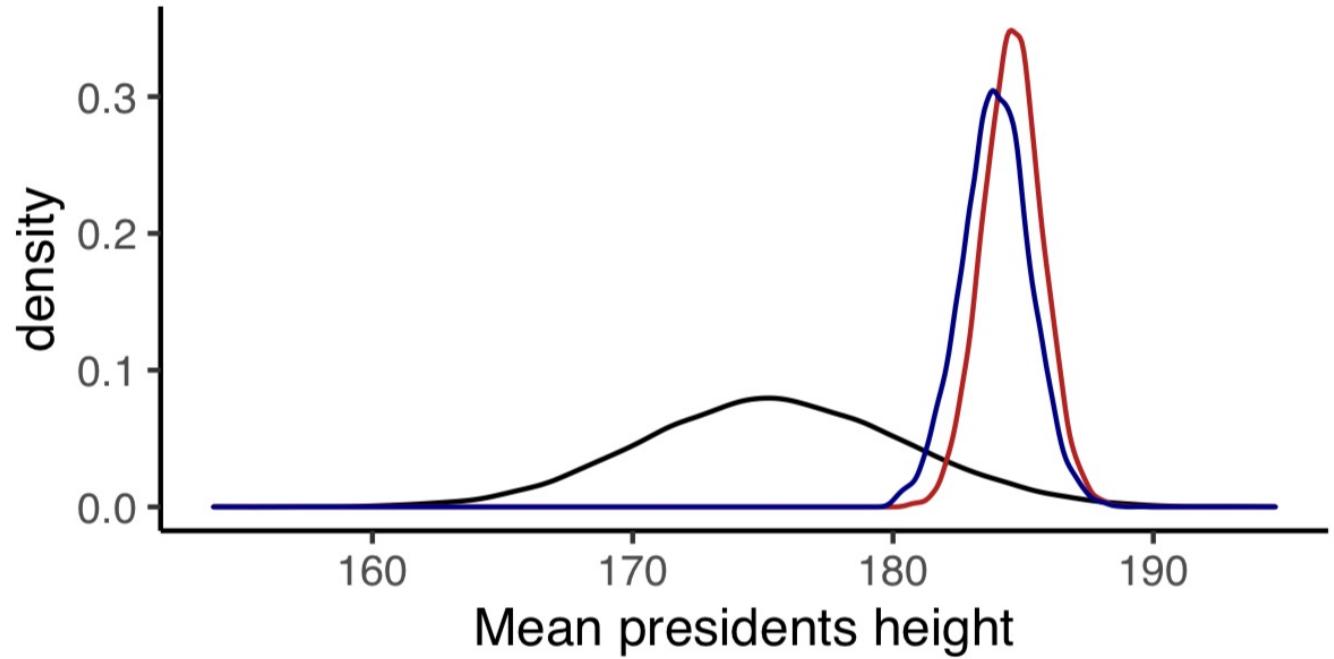
- Predict 4000 new heights: `set.seed(1); rnorm(4000, Mean, SD) %>% sd() ## [1] 4.1`
- This is not ideal, as it does not take into account the fact that our estimates of μ and σ , calculated from limited data, are not precisely true for the full population of US presidents.
- **on top of the randomness that comes from our imperfect knowledge of the mechanisms that generate human heights we must model the uncertainty that comes from imperfect data.**
- we will end up not with point estimates of μ and σ , but probability distributions for μ and σ values. we are not predicting new heights from a single fixed normal distribution, but rather from many different distributions, each is parametrized by a plausible combination of μ and σ values.
 - `pres_m <- brm(data = heights, formula = value ~ 1, file = "models/pres_m")`
 - `posterior_sample <- posterior_samples(pres_m)`
 - `set.seed(1)`
 - `rnorm(4000, posterior_sample$b_Intercept, posterior_sample$sigma) %>% sd() ## [1] 4.546833`
 - the SD is slightly larger (by 0.44 cm), reflecting the extra uncertainty. the model pools both kinds of uncertainties, which we cannot re-separate.



³⁹ `geom_contour()` visualizes 3D surfaces in 2D. To specify a valid surface, the data must contain x, y, and z coordinates, and each unique combination of x and y can appear once.



prior (black), likelihood(red),
and posterior (blue)



Even with only 11 datapoints the prior is so much wider than likelihood that it has minimal influence on the posterior.

Markov Chain Monte Carlo

- we draw a random sample from the posterior
- then we work with this sample numerically
- this is an alternative to the Bayes theorem where we are supposed to test every combination of parameter values against the data. with 10+ parameters, the number of combinations becomes too big for even grid approximation.

MCMC combines two properties: Monte–Carlo and *Markov chain*

- **Monte–Carlo** is the practice of estimating the properties of a distribution by examining random samples from the distribution.
 - to find the mean of a normal distribution, draw a large number of random samples from a normal distribution, and calculate the sample mean of those.
- The **Markov chain** is the idea that the random samples are generated by a special sequential process. Each random sample is used as a stepping stone to generate the next random sample (hence the *chain*).
 - while each new sample depends on the one before it, new samples do *not* depend on any samples before the previous one (this is the “**Markov**” property).

- Q: the mean test score. scores are normal, $sd = 15$. data: 1 score is 100.
- MCMC draws samples from the posterior, which represents the probability of each possible value of the population mean given this single observation.
- start with a plausible random initial value. MCMC is then used to produce a chain of new samples from this initial guess.
- Each new sample is produced by two simple steps:
 1. a *proposal* for the new sample is created by adding a small random perturbation to the most recent sample;
 2. this new proposal is either accepted as the new sample, or rejected (in which case the old sample is retained).

A simple introduction to Markov Chain Monte–Carlo sampling

Don van Ravenzwaaij^{1,2} · Pete Cassey² · Scott D. Brown²

```
samples = numeric(500)                                # 500 samples.  
samples[1]=110←                                     # The initial guess  
for (i in 2:500)  
{  
  proposal = samples[i-1] + rnorm (1, 0, 5) # Proposal value    proposals from rnorm(1, 0, 5)  
  if ((dnorm (proposal, 100, 15) / dnorm (samples[i-1], 100, 15)) > runif (1))  
    samples[i] = proposal                      # Accept proposal  
  else (samples[i] = samples[i-1])              # Reject proposal  
}
```

initial random value

if the normal density is higher at proposal than the current position of the chain, then the ratio > 1 , otherwise it can be larger or smaller than a random number between 0 and 1. If its smaller than..., then the next step is identical to current value.

1. Begin with a plausible *starting value*; 110.
2. Generate a new proposal by taking the last sample (110) and adding random noise, generated from a *proposal distribution*, which should be symmetric and centered on zero ($N(0, 5)$). the new proposal is $110 + \text{a random sample from } N(0, 5)$. Suppose this results in a proposal of 108.
3. Compare the height of the posterior at the value of the proposal against the height at the most recent sample. Since the target distribution is normal with mean 100 (the value of the single observation) and standard deviation 15, this means comparing $N(100|108, 15)$ against $N(100|110, 15)$. $N(\mu|x, \sigma)$ indicates the probability of value μ given the data x and standard deviation σ .
4. If the new proposal has a higher posterior value than the most recent sample, then accept the new proposal. If the new proposal has a lower posterior value than the most recent sample, then randomly choose to accept or reject the new proposal, with a probability equal to the height of both posterior values. For example, if the posterior at the new proposal value is one-fifth as high as the posterior of the most recent sample, then accept the new proposal with 20 % probability.
5. If the new proposal is accepted, it becomes the next sample in the MCMC chain, otherwise the next sample is just a copy of the most recent sample.
6. return to step 2.

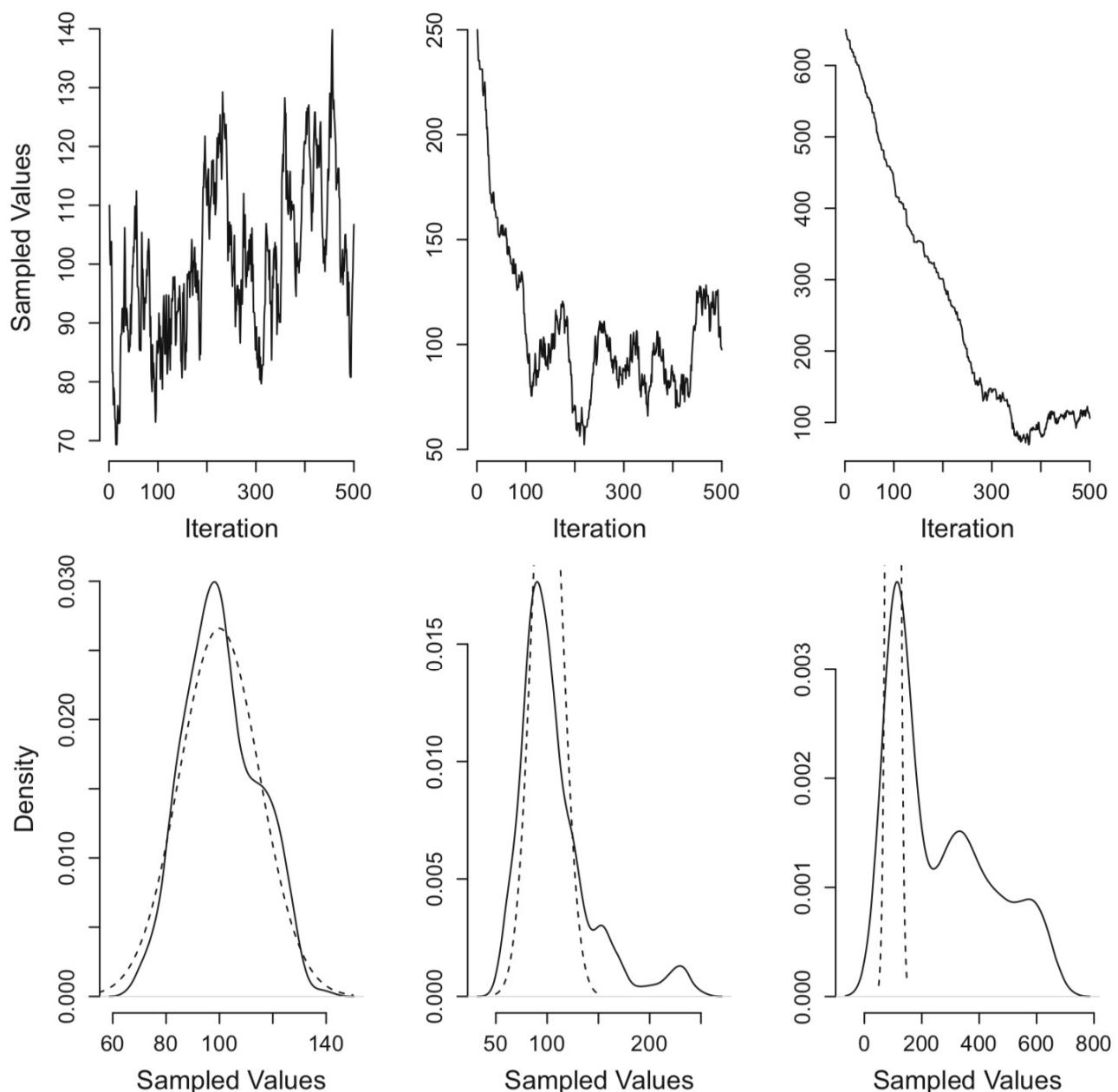


Fig. 1 A simple example of MCMC. *Left column:* A sampling chain starting from a good starting value, the mode of the true distribution. *Middle column:* A sampling chain starting from a starting value in the tails of the true distribution. *Right*

column: A sampling chain starting from a value far from the true distribution. *Top row:* Markov chain. *Bottom row:* sample density. The analytical (true) distribution is indicated by the dashed line

In theory, any symmetric distribution would have worked just as well, but in practice the choice of proposal distribution can greatly influence the performance of the sampler.

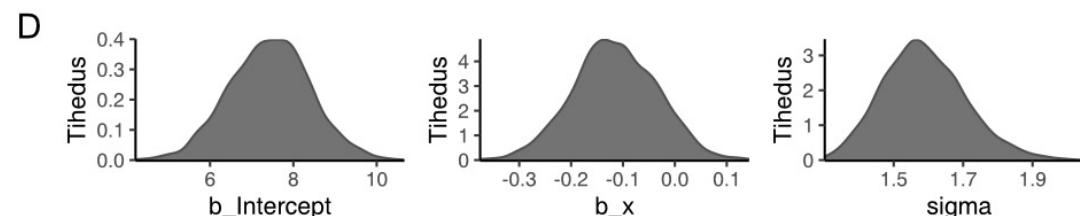
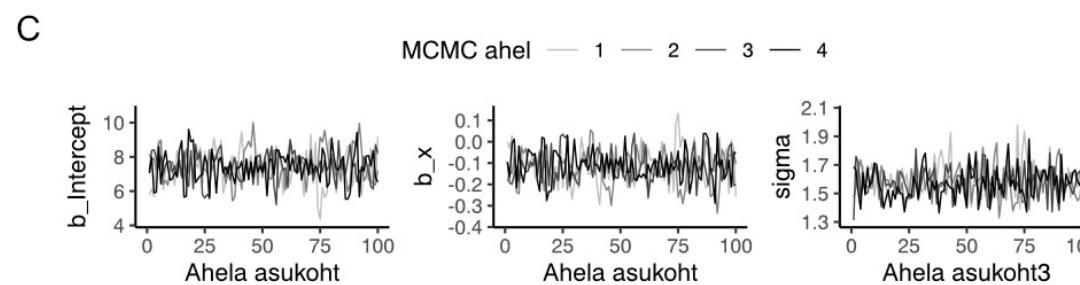
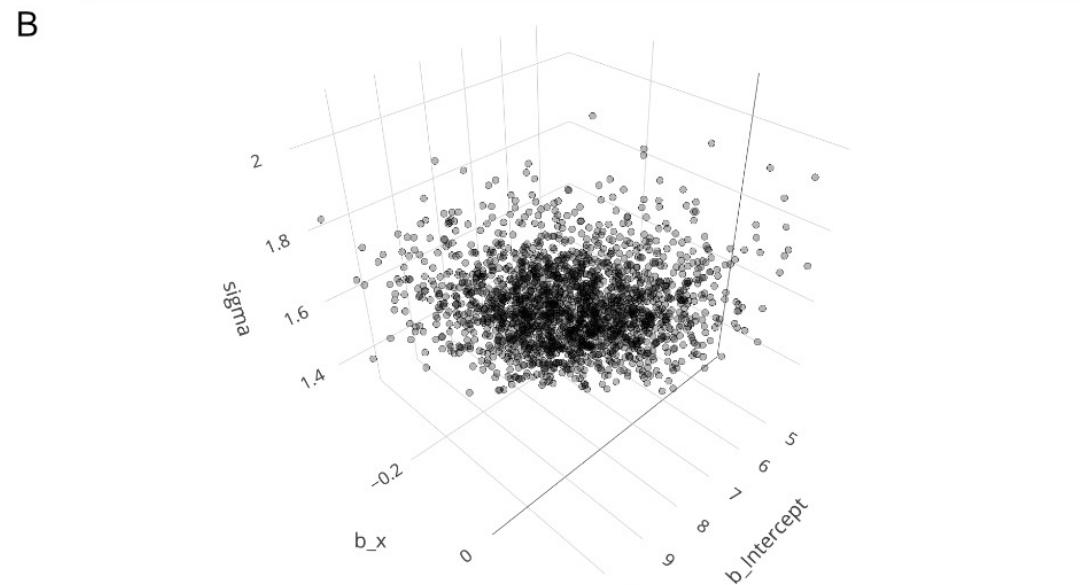
The width of the proposal distribution is called a *tuning parameter*. The fact that the practical performance of the sampler can depend on the value of the tuning parameter is a limitation of the Metropolis–Hastings algorithm, although there are augmented methods that remedy the problem. For example, “auto-tuning” algorithms that adapt the width of the proposal distribution to the nature of the data and distribution.

the fact that initial samples should be ignored as they might be very wrong, deals with a problem known as *convergence and burn-in*.

- use multiple chains to run the sampling many times with different starting values (e.g. with starting values sampled from the prior distribution).
- Differences between the distributions of samples from different chains can indicate problems with burn-in and convergence.
- Another element of the solution is to remove the early samples: those samples from the non-stationary parts of the chain.

A

b_Intercept	b_x	sigma
6.694	-0.05	1.614
6.247	-0.06	1.404
...
6.888	-0.084	1.618



Bayesian estimation

- uncertain knowledge + uncertain data → uncertain knowledge
- prior distribution + likelihood distribution → posterior distribution
- data is fixed, parameter value is estimated
- posterior probability contains all model-based answers about parameter value
- the goal is to honestly propagate uncertainty from data to conclusions

Frequentist vs bayes

freq

- $\Pr = K/N$
- data generating process must be assumed, in order to model the sampling distribution
- goal is action: to fix type I errors
- only data speaks
- models not generative

bayes

- epistemic Pr. - randomness is a property of information, not the world.
- data generating process is not an assumption, but a testable hypothesis
- goal is epistemic: inference of plausible parameter values
- prior knowledge is updated on data
- generative models

frequentist H_0 testing

- sampling distribution (H_0) → p value → if $p < \alpha$, then reject H_0 → fix type I error frequency at significance level (α)
 - type I error freq – nr of rejected true H_0 -s / nr of all tested true H_0 -s
 - sampling distribution – the probability distribution of sample statistics (means, ES-s, etc.) over Inf nr of simulated samples, given that H_0 is true. Data varies, parameters are fixed.
 - p value – the probability of seeing your data or more extreme data given that H_0 is true
- controlling of type I error frequency requires **stopping rules & adjustment for multiple testing.**
 - An important consequence is conservatism, particularly at the early stages of a trial, where the degree of conservatism depends on the stopping rule chosen.
 - if we did an overview of all randomized trials of a given question, and we knew the results of all trials, the type I error of any individual trial would be of little interest.
- To control for multiple testing you need the nr of tests in the family of tests → **pre-registration** of experiment/analysis.

Freq vs. Bayes interpretation

- In frequentist statistics it is usually the raw effect sizes and the p values that get interpreted.
- In Bayesian statistics, they are the shrunk effect sizes and credible intervals.
- In frequentist inference usually a single model is interpreted,
- while in Bayesian inference often several models are interpreted together.
- In frequentist inference the statistically non-significant effects are often not interpreted at all (although this is not what frequentist theory tells you to do)
- in Bayesian inference all effects are interpreted together (*the law of total information* is a law of logic -- you do not want to break these!).

frequentist	Bayesian
Pr = long run relative frequency	Pr = a measure of belief
Parameters/hypotheses do not have Pr	Parameters/hypotheses have Pr
$P(\text{data} \mid H_0)$	$P(H_1 \mid \text{data})$
Algorithms may work poorly for small samples	N=1 is fine
Stopping rules apply	you can peak at data as much as you like
Regularizing algorithms	priors
P values, confidence intervals	posteriors are summarized by credible intervals
Pre-specified modelling	Modelling choices are informed on data
procedures to alleviate the multiple testing problem (Bonferroni, FDR)	This is achieved by model correspondence to data generating mechanism & priors
Single model per effect	Model ensembles
H_0 testing	More flexible hypothesis testing
Central limit theorem applies: if N is large enough, normal models work well.	CLT does not apply, one should specify a likelihood function that models the data generating mechanism
independence assumption leads to randomized designs	Exchangeability assumption is more relaxed, allows correlated data structures.
Hierarchical models work less well	Multilevel (hierarchical) models are the default approach
Some models are generative	All models are generative

presumably you run models for a reason

- prediction of new data (future data)
- drawing scientific inferences from model fit (involves counterfactual prediction from the model)

Füüsilise tervise aspektidest olid tööturul aktiivsete ja mitteaktiivsete vahel kõige suuremad erinevused energia, liikumisvõime ja tööjõu osas. Tööturul aktiivsed uuritavad hindasid võrreldes mitteaktiivsetega kõrgemaks igapäevaeluks vajavat energiat ($p=0,010$), liikumisvõimet ($p=0,002$) ning tööjõudlust ($p=0,001$). Psühholoogilise heaolu aspektist erinesid kõige enam hinnangud elu nautimise osas, kus tööturul aktiivsed olevad uuritavad hindasid palju kõrgemalt oma võimet elu nautida kui mitteaktiivsed ($p=0,001$). Sotsiaalseid suhteid hindasid kõrgemalt tööturul aktiivsed olevad vastajad ($p=0,0437$) (vt Tabel 10).

RCT: drug vs placebo.

$Gr1 = \text{mean}(\text{value})$; $Gr2 = \text{mean}(\text{value})$; $ES = Gr1 - Gr2$

1. 2 means with error bars (SEM, CI, SD) & statistical significance stars for the ES
2. ES & exact p value - is the effect consistent with sampling error?
3. ES & SEM - a 68% CI
4. ES & 90% CI – over many samples 90% of intervals contain the true value
5. ES & 90% credible interval – your sample contains the true value with 90% probability. The true value can be near the lower bound, or near the upper bound, or anywhere in between.
6. Posterior for ES – full description of models uncertainty about the true value.
7. What is the Pr that $ES >$ minimal clinically interesting ES? That it lies between the values x and z?

Bayesian adaptive trial – update prior knowledge on data until sufficient certainty exists to draw conclusions & recommend action

5. MONITORING OF TRIALS

5.1. *Bayesian Monitoring Criteria*

Ethical considerations place an obligation on organizers to take into account all sources of evidence when considering the continuation of a trial (Pocock and Hughes, 1989), not purely the data from the trial itself. A rational way to do this is to use the trial data to provide the likelihood and have all external evidence concerning the effect of treatment on the main clinical outcome represented by a prior distribution. Their combination into a posterior distribution yields the intervals that can be contrasted with the range of equivalence, which combines evidence regarding the relative differences between the treatments in toxicity, cost and inconvenience.

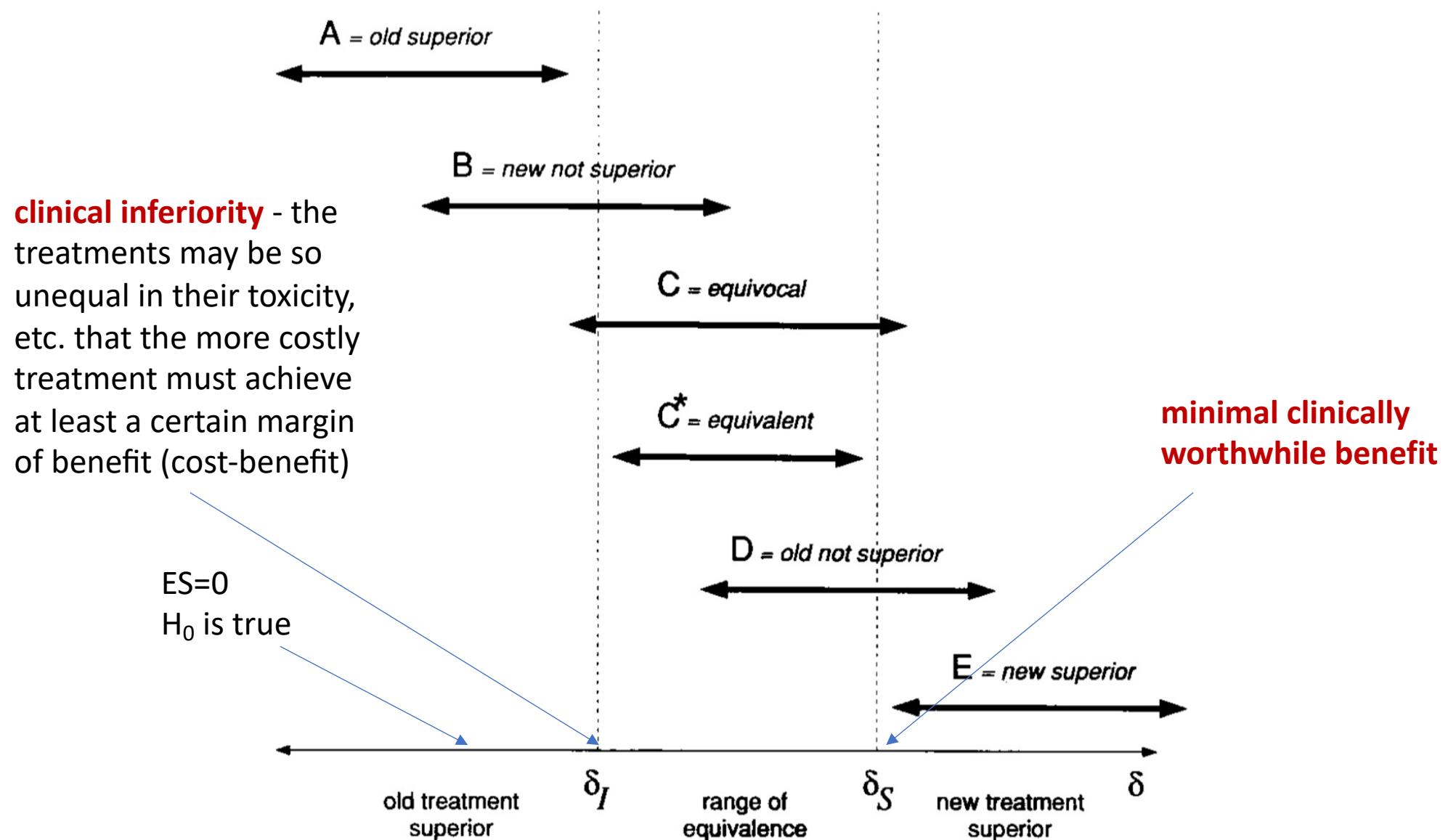


Fig. 4. Possible situations at any point in a trial's progress, derived from superimposing an interval estimate (say 95%) on the range of equivalence (see the text for the relationship between situations A,



Biography

Frank Harrell is a Professor of Biostatistics in the School of Medicine at Vanderbilt University. His research interests include statistical modeling, predictive models and model validation, Bayesian statistics, Bayesian clinical trial design, clinical trial design, analysis, and reporting, statistical computing, statistical graphics, reproducible research, drug development, medical decision making and diagnostic research, health services research, cardiology, COVID-19 therapeutic clinical trial design, and teaching.

Frank Harrell

Professor of Biostatistics
Vanderbilt University



Interests

- Statistical modeling
- Bayesian statistics and study design
- Predictive modeling and model validation

Education

- PhD in Biostatistics, 1979
University of North Carolina
- BS in Mathematics, 1973
University of Alabama in Birmingham

- I started becoming a Bayesian about 1994 because of an influential paper by David Spiegelhalter and because I worked in the same building at Duke University as Don Berry.
- Two other things strongly contributed to my thinking: difficulties explaining p-values and confidence intervals (especially the latter) to clinical researchers, and difficulty of learning group sequential methods in clinical trials. When I talked with Don and learned about the flexibility of the Bayesian approach to clinical trials, and saw Spiegelhalter's embrace of Bayesian methods because of its problem-solving abilities, I was hooked.
- [Don became Bayesian after multiple attempts to teach statistics students the exact definition of a confidence interval. He decided the concept was defective.]

- I was working on clinical trials and started to see that multiplicity adjustments were arbitrary. This started with a clinical trial in which low dose and high dose of a new drug were to be compared to placebo, using an alpha cutoff of 0.03 for each comparison to adjust for multiplicity. The comparison of high dose with placebo resulted in a p-value of 0.04 and the trial was labeled completely “negative” which seemed problematic to me. [Note: the p-value was two-sided and thus didn’t give any special “credit” for the treatment effect coming out in the right direction.]
- I began to see that in biomedical research the typical hypothesis was an artificial construct designed to placate a reviewer who believed that an NIH grant’s specific aims must include null hypotheses. I came to see that questions are more relevant than hypotheses, and estimation was even more important than questions.
- With Bayes, estimation is emphasized. a large number of clinical trials were incorrectly interpreted when $p>0.05$ because the investigators failed to realize that a p-value can only provide evidence against a hypothesis. ...with Bayes it is easy to estimate the probability of similarity of two treatments.
- I listened to clinical trialists debating what should be the primary endpoint in a trial, the co-primary endpoint, the secondary endpoints, co-secondary endpoints, etc. This was all because of their paying attention to alpha-spending. I realized this was all a game.

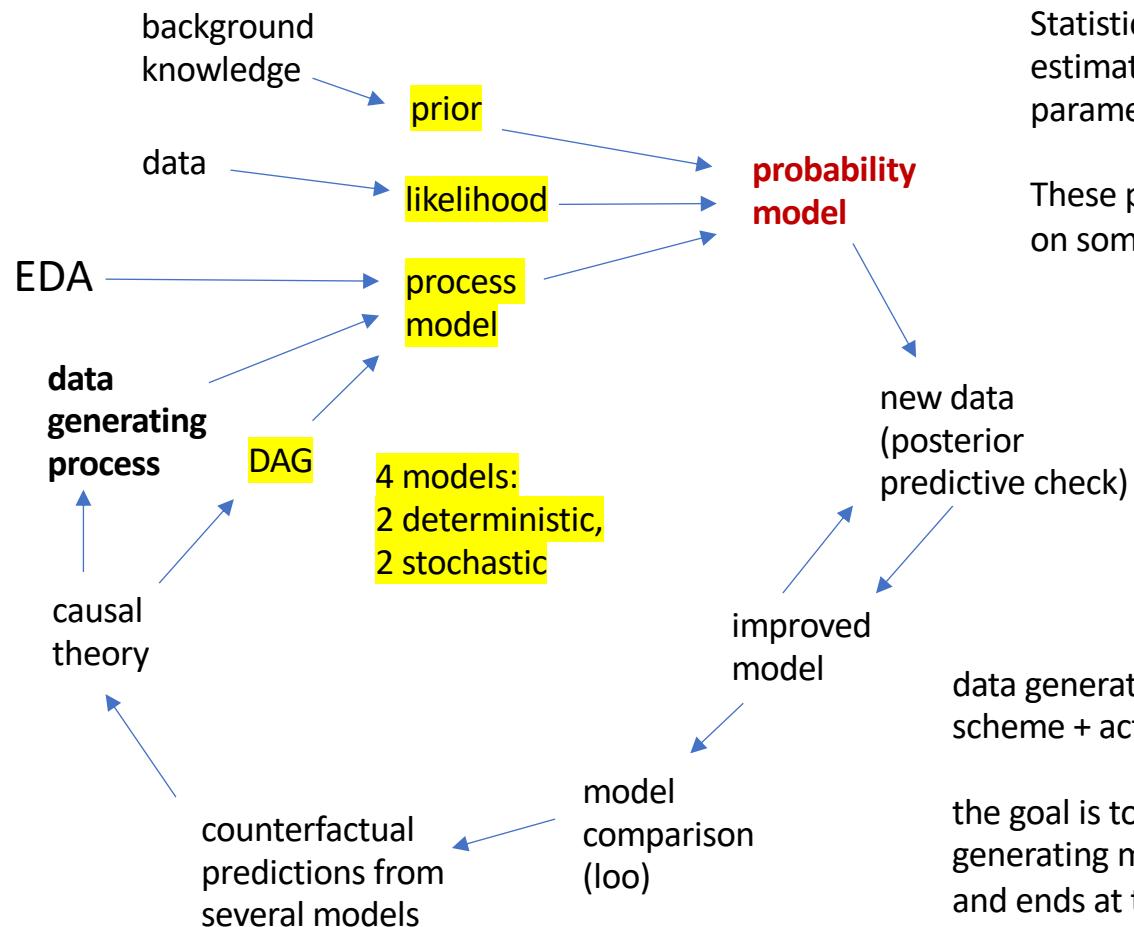
regression workflow

In frequentist statistics the ideal workflow is like this:

1. Decide on the stopping rules and statistical tests that will be used, write a pre-registered data analysis protocol.
2. Do a power analysis to determine sample size N for finding significant effects
3. Collect the sample N (no more, no less).
There can be no peeking at data, unless pre-specified in step 1, until the full sample is collected. The experiment is randomized and blinded.
4. Run the pre-specified statistical test on pre-specified variables, which have been normalized and transformed according to the pre-specified protocol.
5. Check the test assumptions (normality, linearity, constant variance, etc.) by running additional tests on your data and on the fitted model.
6. If assumptions in step 5 are satisfied, interpret the p value.

- In Bayesian statistics instead of a power analysis we simulate realistic-looking data and analyse these, to see whether the models and algorithms to fit them work well enough for these kind of data.
 - This will double as a power analysis, as one can find an optimal N that gives sensible estimates.
- we do not pre-specify models, as we will build several bespoke models depending on the data that we get, so we can compare the models and their fits, and try to learn from them as a group.
- Instead of formally checking modelling assumptions, we generate simulated data from the fitted models and compare with actual data, to see, which aspects of our data are well represented in the model structure, and which ones are not.
- Bayesian modelling is more flexible, both in terms of model structures and in how we work with the models, in that it tries to mirror the scientific process, where we gradually learn from data, and from our mistakes.

Bayesian cycle



Statistical inference is a set of operations on data that yield estimates and uncertainty statements about predictions and parameters of some underlying process or population.

These probabilistic uncertainty statements are derived based on some assumed probability model for observed data.

data generating process: scientific hypothesis + data collection scheme + actual process of data collection/measurement

the goal is to infer stuff about the scientific part of the data generating mechanism based on data & model. Everything starts and ends at the data generating process.

Scientific hypotheses are causal, thus modelling has to be causal, but causal knowledge is not in the data

A simplified Bayesian workflow:

1. model data as likelihood, by using an appropriate probability distribution that reflects the data generating mechanism.
2. model additional information as prior, by using a probability distribution that best allows to describe your prior beliefs.
3. use a Bayesian (MCMC) procedure to get a posterior.
4. compare the model fits, and the posteriors that we got from different models
5. As needed, modify your models to achieve better fit and more realistic model-generated data. Then repeat the cycle, until you are satisfied.
6. interpret the posteriors in scientific context, make counterfactual predictions from the fitted model.



ORIGINAL ARTICLE



Causal variation modelling identifies large inter- and intra-regional disparities in physical therapy offered to hip fracture patients in Estonia

Pärt Prommik^{a,b,c*} , Ülo Maiväli^{d*}, Helgi Kolk^{a,b†} and Aare Märtsom^{a,b‡}

^aDepartment of Traumatology and Orthopaedics, University of Tartu, Tartu, Estonia; ^bTraumatology and Orthopaedics Clinic, Tartu University Hospital, Tartu, Estonia; ^cInstitute of Sport Sciences and Physiotherapy, Tartu, Estonia; ^dInstitute of Technology, University of Tartu, Tartu, Estonia

ABSTRACT

Purpose: An essential measure of hip fracture (HF) rehabilitation, the amount of physical therapy (PT) used per patient, has been severely understudied. This study (1) evaluates post-acute PT use after HF in Estonia, (2) presents causal variation modelling for examining inter- and intra-regional disparities, and (3) analyses its temporal trends.

Materials and methods: This retrospective cohort study used validated population-wide health data, including patients aged ≥ 50 years, with an index HF diagnosed between January 2009 and September 2017. Patients' 6-month PT use was analysed and reported separately for acute and post-acute phases.

Results: While most of the included 11,461 patients received acute rehabilitation, only 40% of them received post-acute PT by a median of 6 h. Analyses based on measures of central tendency revealed 2.5 to 2.6-fold inter-regional differences in HF post-acute rehabilitation. Variation modelling additionally detected intra-regional disparities, showing imbalances in the fairness of allocating local rehabilitation resources between a county's patients.

Conclusions: This study demonstrates the advantages of causal variation modelling for identifying inter- and intra-regional disparities in rehabilitation. The analyses revealed persisting large multi-level disparities and accompanying overall inaccessibility of PT in HF rehabilitation in Estonia, showing an urgent need for system-wide improvements.

ARTICLE HISTORY

Received 16 December 2020

Revised 9 April 2021

Accepted 14 April 2021

KEYWORDS

Hip fracture; rehabilitation; research methodology; causal directed acyclic graphs; variation modelling; zero-inflation

describe the process of getting/not getting post-acute physiotherapy (PT)

hip fracture → acute care (incl. possible PT) → post-acute care w. or without PT

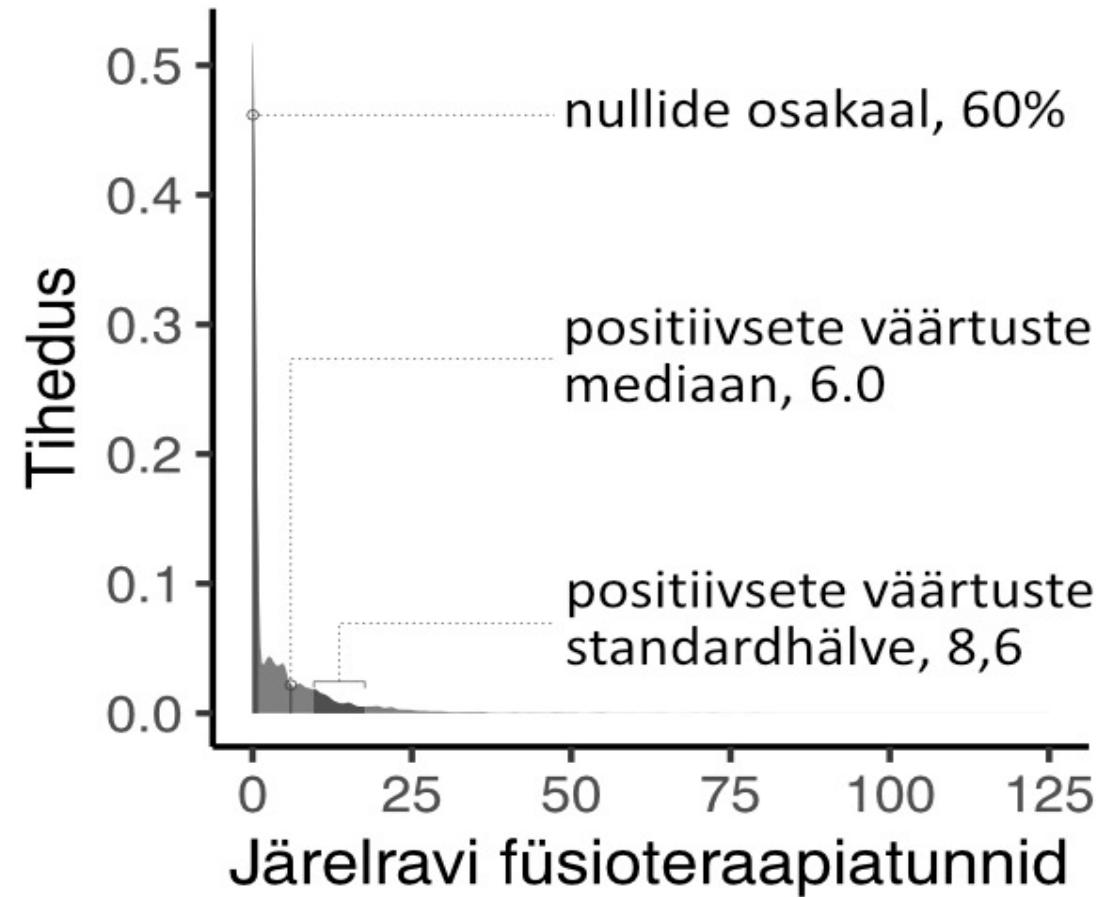
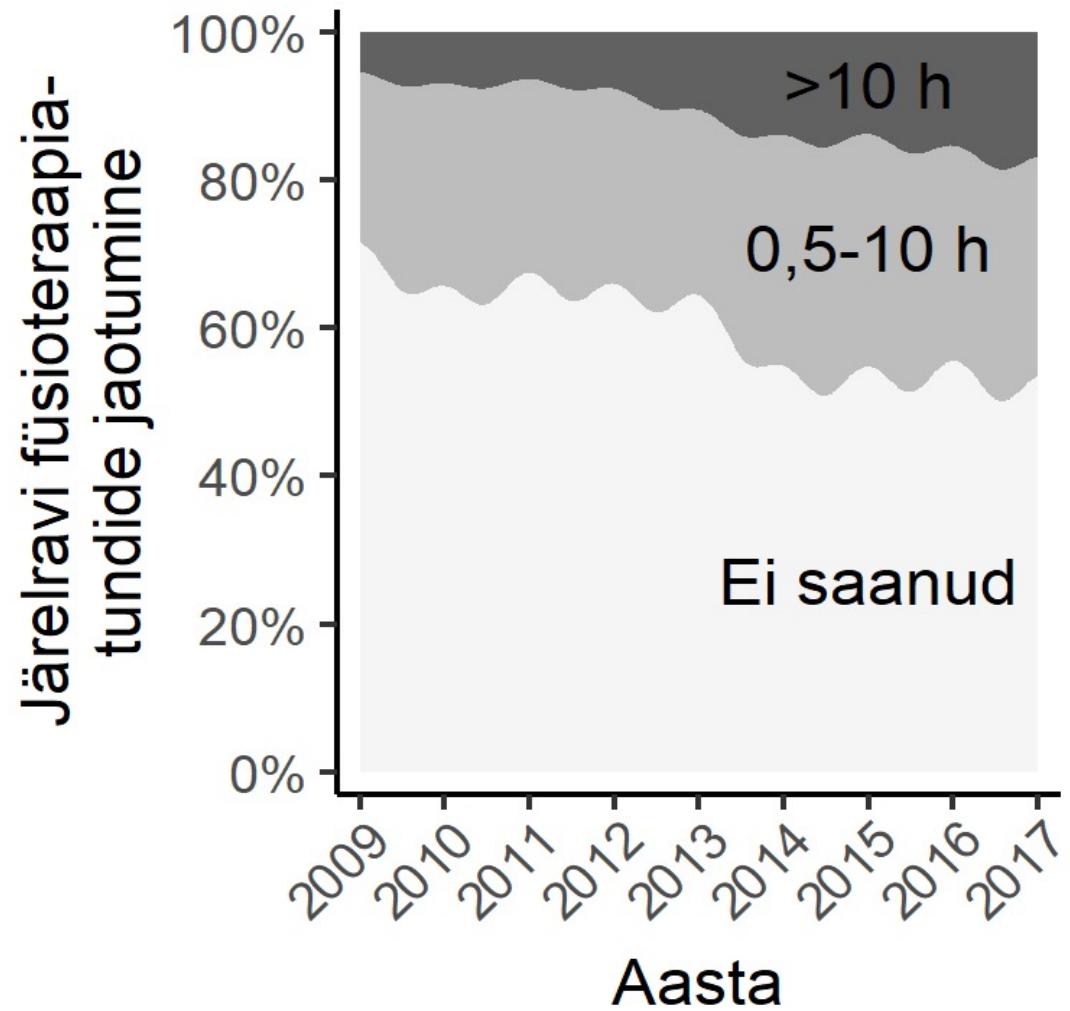
Andmed ja uurimisküsimus

- Populatsioonipõhised andmed koguti Eesti Haigekassa andmebaasist.
 - 50+ aastased, kellel diagnoositi jaanuar 2009 – september 2017 reieluu proksimaalse osa murd.
- Andmed sisaldasid füsioteraapia kasutamist aktiiv- ja järelravis. aktiivravis haige hospitaliseeritakse, opereeritakse ja seejärel alustatakse füsioteraapiaga.
- Võrdlemisi lühikesele aktiivravile järgneb järelravi, kus peamine fookus on kehalise võimekuse taastamisel. Just järelravi ajal oli füsioteraapia haigetele halvasti kätesaadav.

1. EDA annab aimu nii andmete kvaliteedi, kui formaalse mudeldamise võimalike tulemuste kohta.

- Lisaks aitab parem arusaam andmetest meil oma mudeleid paremini spetsifitseerida, näiteks valida ühe- ja mitmetasemeliste mudelite vahel.
- Me analüüsime 11 461 haige taastusravi.
- aktiivravi ajal ei saanud 24% haigetest füsioteraapiat ja et järelravi käigus ei saanud 60% haigetest füsioteraapiat ning ülejäänud said seda võrdlemisi väikeses koguses (mediaan 6 tundi, 25- ja 75-protsentiil: 3, 11).
- Järelravi ajal füsioteraapiat mittesaanute osakaal vähenes üheksa-aastase perioodi jooksul ja seda > 10 tunni saanute osakaal suurennes.

Lisaks esineb maakondade vaheline erinevus taastusravis ja selle trendides, mis viitab, et mitmetasemelised mudelid võiksid hästi sobida nende haigete järelravi ajal saadud füsioteraapia muutuste analüüsiks. Mudeldamise seisukohast on oluline ka peamise tulemuslikkuse mõõdiku – järelravi ajal saadud füsioteraapiatundide arvu – jaotus, mis sisaldab palju nulle, näidates mitmeosalise regressioonanalüüsijajalikkust.



2. Tõepärafunktsiooni leidmine

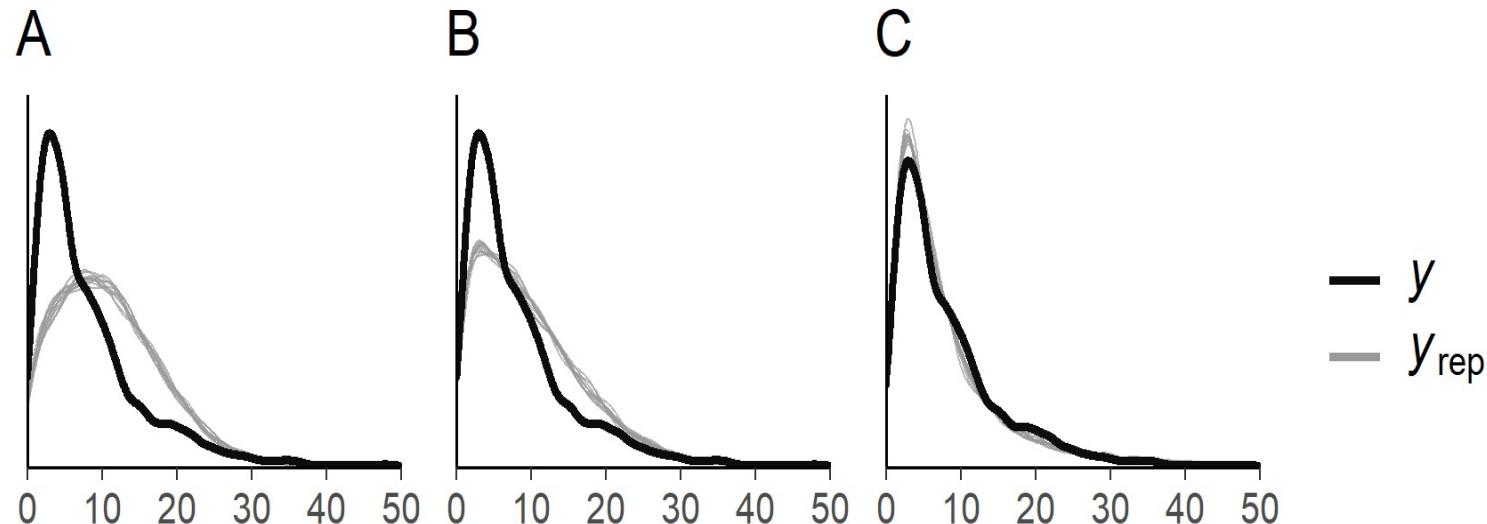
- Tõepärafunktsioon mudeldab sõltuva tunnuse (y) väärtsi tingimuslikena x -muutujate (prediktorite) väärustest.
- Kuna meil on prediktoriks nullide all lookas ja tugevalt paremale kiivas pidev tunnus, siis analüüsime ajalisi trende kolmes osas:
 - (i) füsioteraapiat saanute osakaal,
 - (ii) saadud füsioteraapiatundide arv ja
 - (iii) saadud füsioteraapiatudide varieeruvus patsientide vahel.
- Kuna kõik bayesaanlikud mudelid on generatiivsed, aitab tõepärafunktsiooni valida *posterior predictive check*.
 - Kui võrdlus näitab halba kokkulangemist, siis võime muuta tõepärafunktsiooni, sõltumatuid tunnuseid, nendevahelisi seoseid või lisada mudelisse konditsionaalseid varieeruvusi. Sõltuvalt analüüsi eesmärgist, võib meid huvitada andmejaotuste sobivuse erinevad aspektid. Näiteks kui oleme huvitatud keskväärtustest, siis võrdleme *in silico* valimite keskväärtusi empiirilise valimi keskväärtusega, kui meid huvitavad varieeruvused või näiteks maksimumväärtused, siis võrdleme hoopis neid.

- 1. kuidas füsioteraapiat saanute osakaal muutus järelravi ajal? muutsumise sõltuva tunnuse binaarseks („Ei saanud”, „Sai“) ja mudeldame Bernoulli jaotusega.
 - *terapia_jarelravi_binaarne ~ aasta, family = „bernoulli“, link = „logit“*
- 2. järelravi ajal saadud füsioteraapia maht. proovisime kolme tõepärafunktsiooni: normaaljaotus, kallutatud normaaljaotus ja lognormaaljaotus. Lisaks täiendame kõiki kolme mudeliteid selliselt, et nad üheaegselt mudeldaksid sõltuva tunnuse varieeruvuse sõltuvust sõltumatutest tunnusest. Sellised mudelid annavad üheaegselt hinnangu sõltuvalt tunnusele ja ka selle varieeruvusele igal sõltumatu tunnuse väärthusel. Vastavad mudelid spetsifitseeritakse brms-is järgmiselt:
 - *brm(bf(terapia_jarelravi ~ aasta, sigma ~ aasta), family = „gaussian“, data = andmed)*
 - *brm(bf(terapia_jarelravi ~ aasta, sigma ~ aasta), family = „skew_normal“, data = andmed)*
 - *brm(bf(terapia_jarelravi ~ aasta, sigma ~ aasta), family = „lognormal“, data = andmed)*

brm(bf(terapia_jarelravi ~ aasta, sigma ~ aasta), family = „gaussian“, data = andmed)

brm(bf(terapia_jarelravi ~ aasta, sigma ~ aasta), family = „skew_normal“, data = andmed)

brm(bf(terapia_jarelravi ~ aasta, sigma ~ aasta), family = „lognormal“, data = andmed)



3. Protsessimudelite ehitamine

- Ajaliste trendide spetsifitseerimisel arvestasime:
 - (i) haige taastumise seisukohast ei ole vahet, millises ravifaasis taastusravi läbi viakse.
 - (ii) Andmestruktuuri hierarhilisus: erinevates maakondades pakutakse taastusravi suhteliselt iseseisvalt.
 - (iii) haigete taastusravi ja selle ajalised muutused erinesid maakonniti.
 - kuus erineva keerukusega Bernoulli- (ravi saajad) ja lognormaalsel jaotusel (terapiatundide arv) põhinevat mudelit, millest viimane sisaldas ainult positiivseid väärthusi.
 1. $\text{terapia_jarelravi} \sim 1$
 2. $\text{terapia_jarelravi} \sim \text{aasta}$
 3. $\text{terapia_jarelravi} \sim \text{aasta} + \text{terapia_aktiivravi}$
 4. $\text{terapia_jarelravi} \sim \text{aasta} + \text{terapia_aktiivravi} + \text{maakond}$
 5. $\text{terapia_jarelravi} \sim \text{aasta} + \text{terapia_aktiivravi} + (1 / \text{maakond})$
 6. $\text{terapia_jarelravi} \sim \text{aasta} + \text{terapia_aktiivravi} + (\text{aasta} / \text{maakond})$
- Lognormaalsed mudelid mudeldasid ka SDd, kasutades samu x-tunnuseid

leave-one-out (loo) ristvalideerimise võrdlus võimaldas valida optimaalse keerukuse-ennustusvõime suhtega mudelid (loo annab hinnangu mudeli valimivälise sobivuse kohta andmetega).

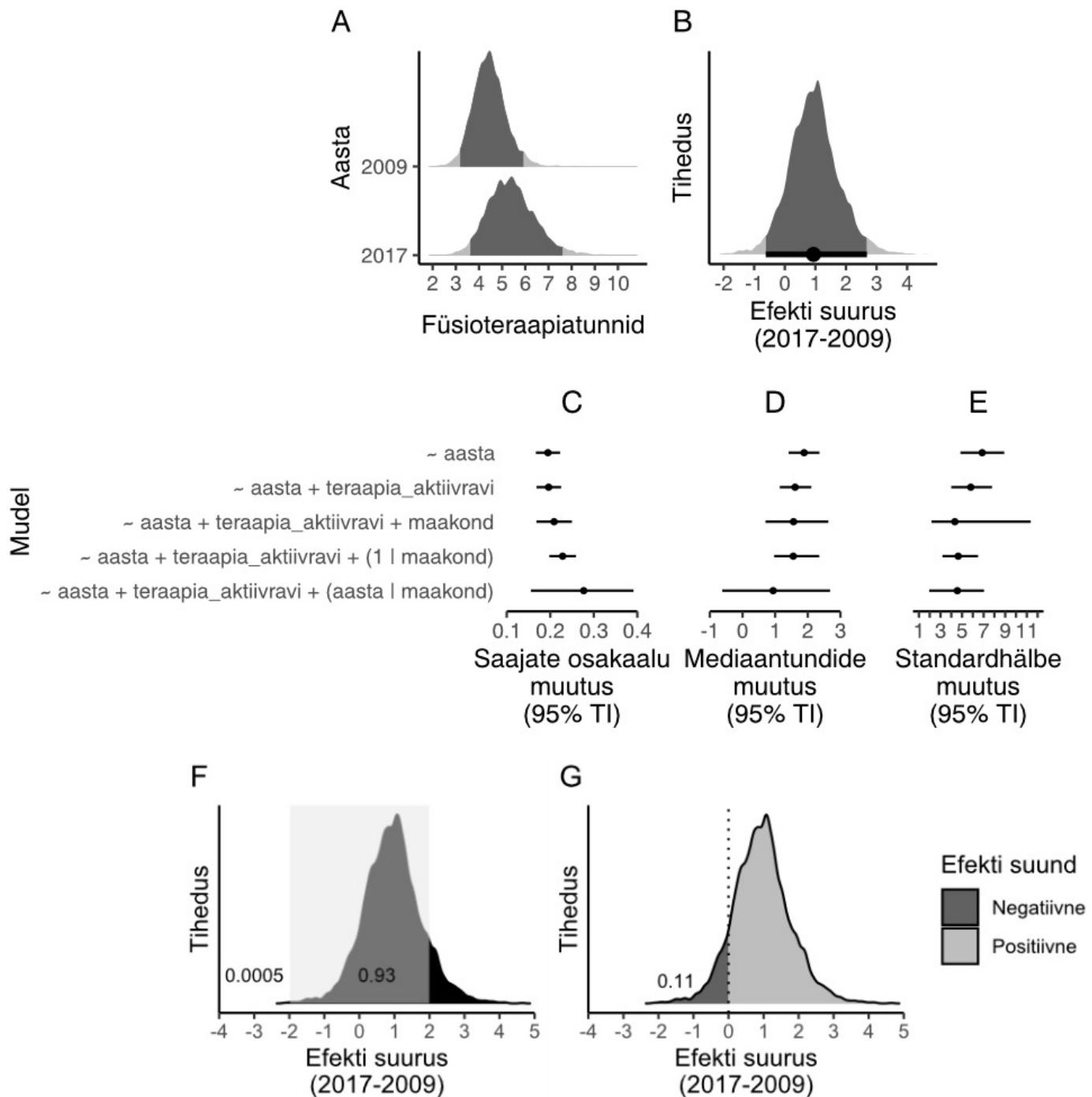
kui $se_diff > 3 \times elpd_diff$, siis mudel on eelmisest oluliselt halvem.

saame ka SD mudeldamise mõju mudeli headusele. SD mittemudeldamine halvendas mudeli sobivust ($elpd_diff -193.7$ ja $se_diff 19.1$). Seda võib tõlgendada kinnitusena, et meie arusaam andmete hierarhilisest struktuurist on adekvaatne.

Mudel	Bernoulli		Lognormaalne	
	elpd_diff	se_diff	elpd_diff	se_diff
$\sim aasta + teraapia_aktiivravi + (aasta / maakond)^*$	0.0	0.0	0.0	0.0
$\sim aasta + teraapia_aktiivravi + (1 / maakond)^*$	0.0	0.0	-32.9	8.9
$\sim aasta + teraapia_aktiivravi + maakond^*$	-0.1	0.5	-33.3	9.2
$\sim aasta + teraapia_aktiivravi^*$	-492.6	30.9	-308.7	25.7
$\sim aasta^*$	-521.9	31.7	-342.7	27.3
$\sim 1^*$	-611.4	34.1	-376.1	28.3

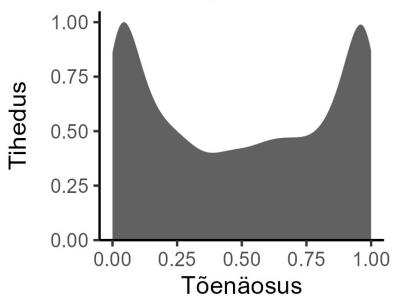
*samu sõltumaid tunnuseid kasutati standardhälbe paralleelseks mudeldamiseks lognormaalsetes mudelites; elpd näitab määratud mudeli sobivust andmetega, ehk mudeli ennustuste täpsust; elpd-diff – oodatud logaritmitud ennustatavate tiheduste erinevus (*expected log-predictive density*); se_diff annab standardhälbe elpd_diff suuruse hinnangu järeljaotusele;

- Me saame järeljaotust kirjeldada kasutades tavapäraseid keskmist iseloomustavaid statistikuid koos tõenäosusintervalliga. Näiteks 2009. aastal said haiged järelravi ajal mediaanina 4,4 tundi [95% TI: 3,2; 5,9] füsioteraapiat ja 2017. aastal 5,3 tundi [95% TI: 3,6; 7,6]
- Teiseks saame järeljaotustest arvutada uusi huvitavaid järeljaotusi. Näiteks taastusravi 9-aastase trendi kirjeldamiseks lahutame ühest järeljaotusest teise, saades efekti suuruse ($ES = Y_{2017} - Y_{2009}$) järeljaotuse ($ES = 1.0$ [95% TI: -0.6; 2.7]).
- Lisaks kasutame järeljaotusi spetsiifiliste hüpoteeside testimiseks.
 - Näiteks võime hüpoteesina defineerida kliiniliselt ebaolulise muutuse trendi, mis on väiksem kui 2 tundi 9 aasta vältel. Kuna ES järeljaotuse tõenäosusmassist jäab 93% selle hüpoteesi poolt defineeritud vahemikku, siis on kliiniliselt olulise positiivse ajalise muutuse tõenäosus 0.07.
 - Alternatiivina arvutame bayesiaanliku p-väärtuse, ehk testime hüpoteesi, mille kohaselt keskmiste ravitundide arv on ajas langev. Selle hüpoteesi tõenäosus on 0.11.
 - Samas, tõenäosus, et trend on ajas kangev ja samas kliinilises mõttes olulise suurusega, on 0,0005. Seega, ehkki meie testid viitavad kokkuvõttes sellele, et mediaan-füsioteraapiatunnid pigem ei kasvanud 9 aasta jooksul kliiniliselt olulisel määral, jätab nad siiski arvestatavat kaalu hüpoteesile, mille kohaselt selline kasv ikkagi toimus.

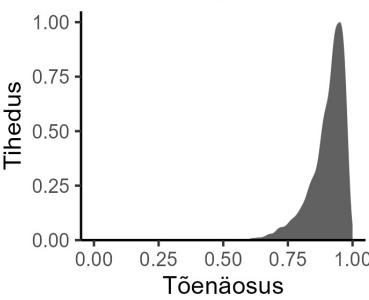


- Kuna eeljaotus võib eelteadmise peegeldajana meid viia valedele radadele, tuleks kontrollida selle mõju järeljaotusele. Selline kontroll näitab ka andmete mõju ulatust määratud mudelile, mida on eriti keerulisemate mudelite korral hea teada. Range eeljaotuse mõju demonstreerimiseks võrdleme kahte versiooni sama protsessimudeliga hierarhilisest Bernoulli mudelist
 - $\text{brm}(\text{terapia_jarelravi_binaarne} \sim \text{aasta} + \text{terapia_aktiivravi} + (\text{aasta} / \text{maakond}), \text{family} = \text{„beroulli”}, \text{link} = \text{„logit”}, \text{data} = \text{andmed})$.
- Ühe määrame brms-i vaike-eeljaotusega ja teisele anname range ja ebamõistliku eeljaotuse, mis ilustab Eesti haigete ravikäsitlust. See ebasiiras eeljaotus ütleb, et me usume tõenäosusega 0.96, et Eestis taastusravi saajate osakaal on >75% ja et me usume vaid 0.0008 tõenäosusega, et see on <50% (vastavalt taastusravi ravijuhistele peaks see olema lähedal 100%-le).
- meie suur andmestik lahustas pea täielikult eeljaotuse mõju: range eeljaotusega mudelis tõusis terapia saajate osakaal keskmiselt vaid 3,6 protsendipunkti võrra [95% TI: -11,1; 19,1].

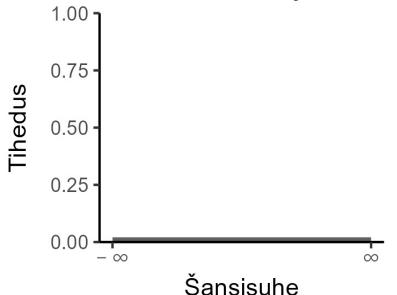
A Vabaliikme eeljaotus



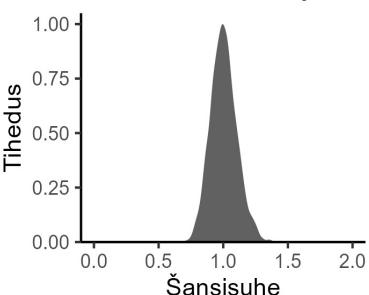
B Vabaliikme eeljaotus



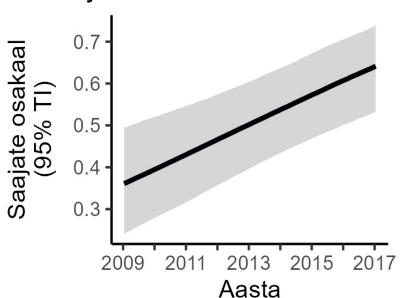
C Tõusukoefitsendi eeljaotus



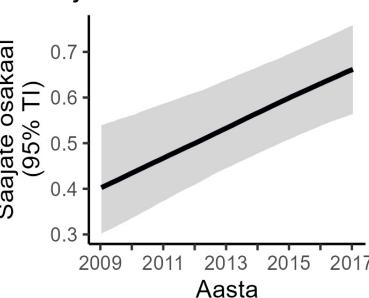
D Tõusukoefitsendi eeljaotus



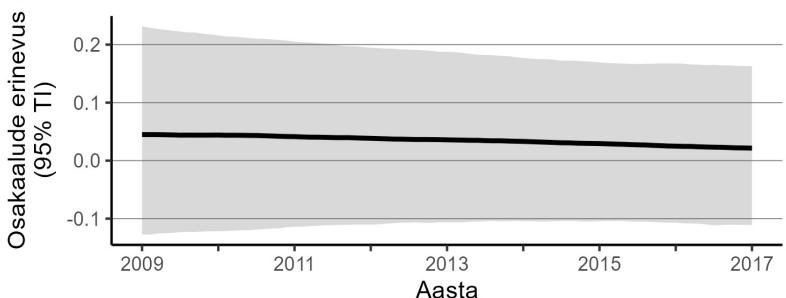
E Saajate osakaalu muutus



F Saajate osakaalu muutus



G Kahe mudeli osakaalu muutuste erinevus



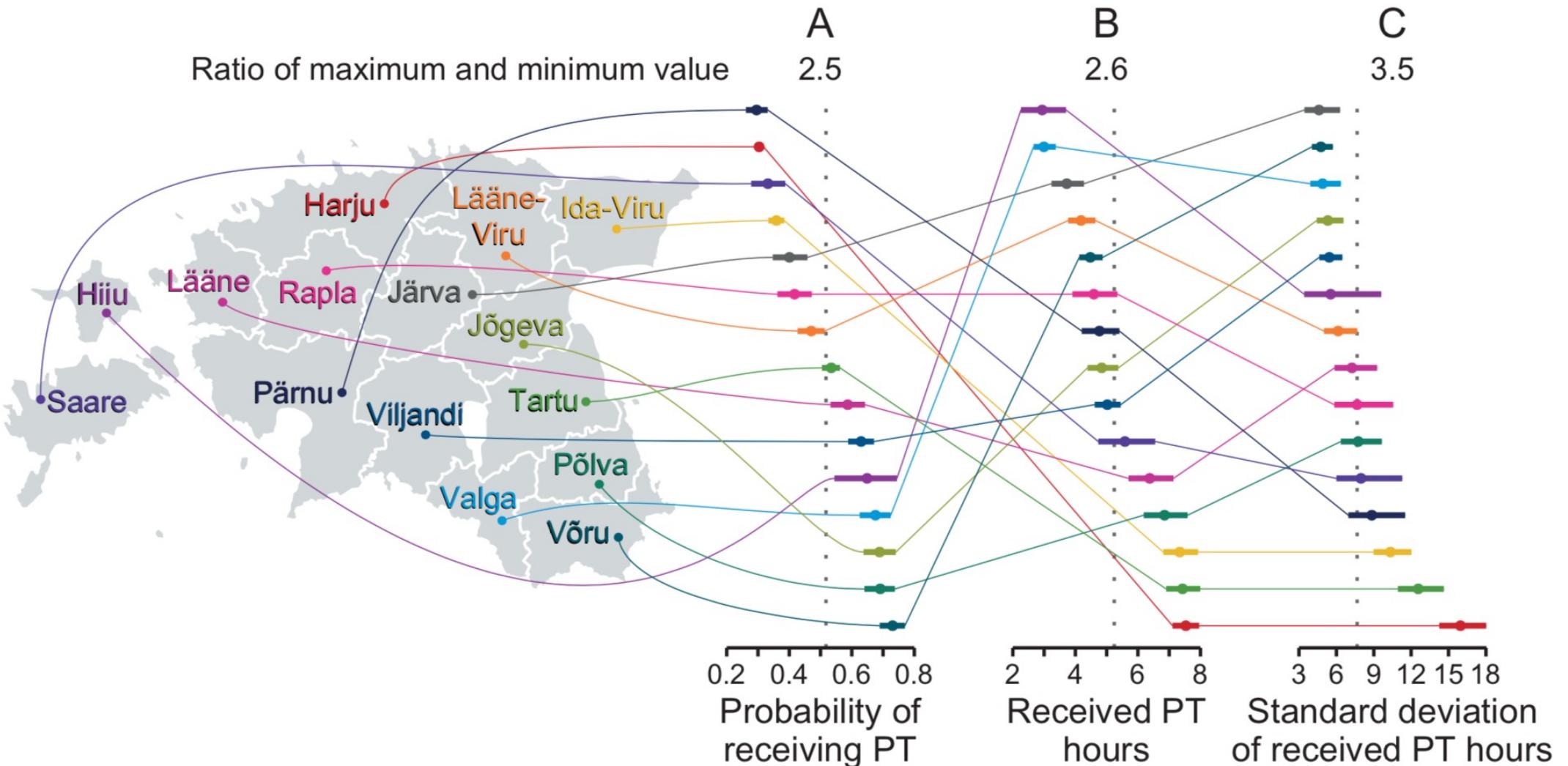


Figure 4. Adjusted inter- (A, B) and intra-regional (C) comparisons of the total received post-acute physical therapy hours. (A) Mean probability for receiving post-acute physical therapy. (B) Median total received post-acute physical therapy hours. (C) The standard deviation of received physical therapy hours. Point estimates are given with 95% credible intervals. Dotted lines give the respective estimates for the whole country. PT: physical therapy.

Mudelite ansamblli põhised konsistentsed ajalised trendid võimaldavad teha järgmisi teaduslikke järeldusi:

- 1) füsioteraapia saajate osakaal suurenemine järelravi ajal 28 protsendipunkti võrra [95% TI: 16; 39], mis annab tunnistust riiklikul tasemel toimunud pingutusest parema järelravi tagamiseks.
- 2) Kuid füsioteraapiat saatnute teraapiatundide mediaan ei kasvanud kuigivõrd ning
- 3) füsioteraapiatundide jagamine patsientide vahel muutus üheksa-aastasel perioodil 1,9 korda ebaühtlasemaks [95% TI: 1,3; 2,6] ja patsiendi tasemel saadud teraapiatundide standardhälve suurenemine 4,6 tunni võrra [95% TI: 2,0; 7,0], mis viitab kohalike ravikeskuste tasemel aset leidva ebaõigluse süvenemisele raviresursi jagamisel patsientide vahel.
- 4) maakondlikud hinnangud näitavad suurt varieeruvust maakondade vahel kõigis kolmes näitajas, ehk seda, et nii lisandununud raviresurss kui selle kasutamise efektiivsus liikus kohalikule tasemele väga ebaühtlaselt. Üldiselt on Lõuna-Eesti tunduvalt paremas seisus kui rannikuga piirnevad maakonnad, eriti Harjumaa, kus ravi saamise ebavõrdsus on väga suur ja ravi mittesaanute osakaal on väga kõrge.

- EDA andis aimduse, milliseid trende võksid regressioonimudelid edasises analüüs is näidata. Modelitest *in silico* genereeritud andmete võrdlemine sõltuva tunnuse tegeliku jaotusega aitas meil valida tõepärafunktsioonid ja mudelite determinismliku osa struktuurid, mis võimaldavad realistlikke mudelipõhiseid ennustusi. Mitme mudeli tulemuste üheskoos vaatlemine andis parema arusaama tegeliku mõju suunast ja suurusest. Eeljaotuse mõju hindamine kinnitas meie tulemuste valiidsust, näidates, et andmete suur hulk lahjendas ka väga tugeva eeljaotuse mõju sisuliselt olematuks.

bayesaanlikul lähenemise puudused

- (i) Puudub sisse töötatud vorm tulemuste avaldamiseks. Siin aitab analüüsikoodi täismahus avaldamine artikli lisana, soovitavalt koos lühikese õpetusega, kuidas seda koodi kasutada ja selles olevaid mudeliobjekte avada.
- (ii) Informatiivsete eeljaotuste defineerimine keerukamatele mudelitele võib osutuda keerukaks. Selle probleemiga tegeletakse aktiivselt.
- (iii) Töö mudelite ansambliga loob ohu mudelite ülemääramiseks ja seega ületõlgendamiseks.
 - Kui me eelistame tõlgendada mudeliteid, mis paremini sobituvad meie valimiandmetega, siis jäab alati oht, et meie valimiandmed valetavad ja me teeme liiga julgeid järelusi nende hea sobivuse põhjal mudelitega, millised me oleme välja valinud just sellesama sobivuse tõttu. Paraku alternatiiv, kus me arvutame läbi ühe mudeli ja elame siis selle mudeli ennustustega hoolimata sellest, kui hästi see mudel generatiivselt töötab või andmete struktuuriga sobitub, ei tundu meile kuigi ahvatlev. Mudelite ülemääramist aitab vältida eeskõige mudeli sobivus andmeid genereeriva protsessiga, eriti mitmetasemeliste mudelite korral. Sellega seoses on just bayesiaanlikus statistikas eriti oluline regressiooni sidumine formaalsete põhjuslike mudelitega (Pearl 2018).
- (iv) retsensendid kipuvad võtma bayesiaanlikku andmeanalüüsni kriitikavabalt.
- (v) mudelid jooksevad aeglaselt.