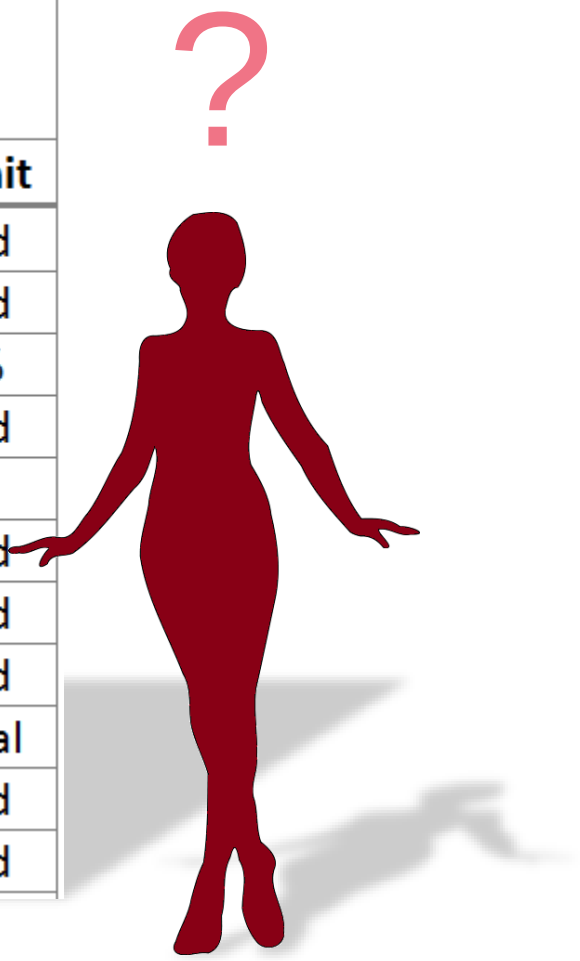
Modeling the Effects of Nutrition with Mixed-Effect Bayesian Network

Jari Turkia
CGI, Next
University of Eastern Finland, School of Computing

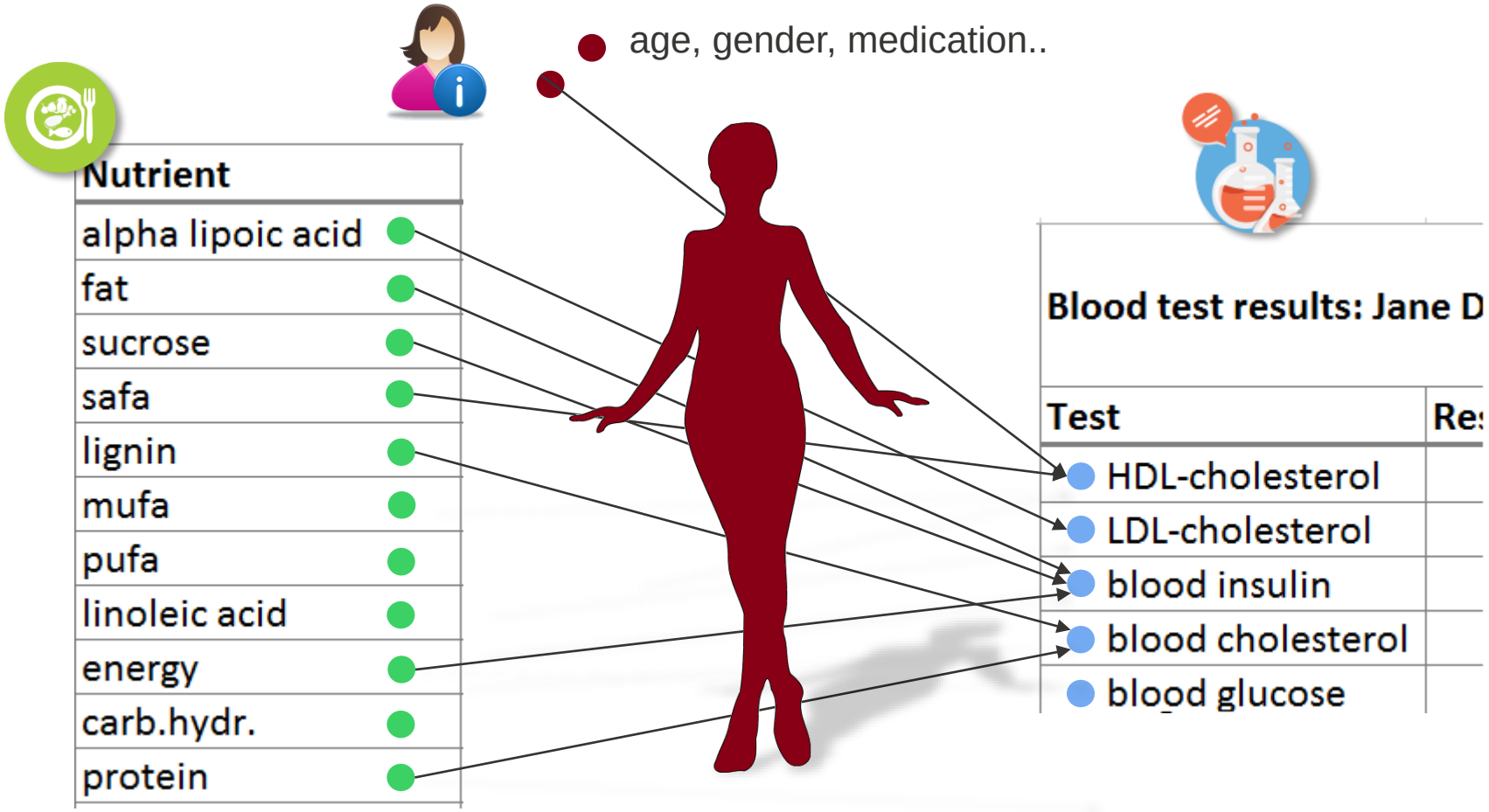

Do eggs raise cholesterol levels?
Why my insulin levels are high?
Can I have more coffee?
What if you could eat in
a way your body works?

General guidelines won't answer to these!

Recommendation: VRN-14/daily: Women 31-60 yrs			
Nutrient	**Lower limit**	**Upper limit**	**Unit**
alpha lipoic acid	36,00	45,00	g/d
fat	59,00	106,00	g/d
sucrose		10,00	E%
safa	23,00	53,00	g/d
lignin			
mufa	12,00	26,00	g/d
pufa	12,00	26,00	g/d
linoleic acid	0,71	1,53	g/d
energy	2110,00	2380,00	kcal
carb.hydr.	237,00	357,00	g/d
protein	53,00	119,00	g/d



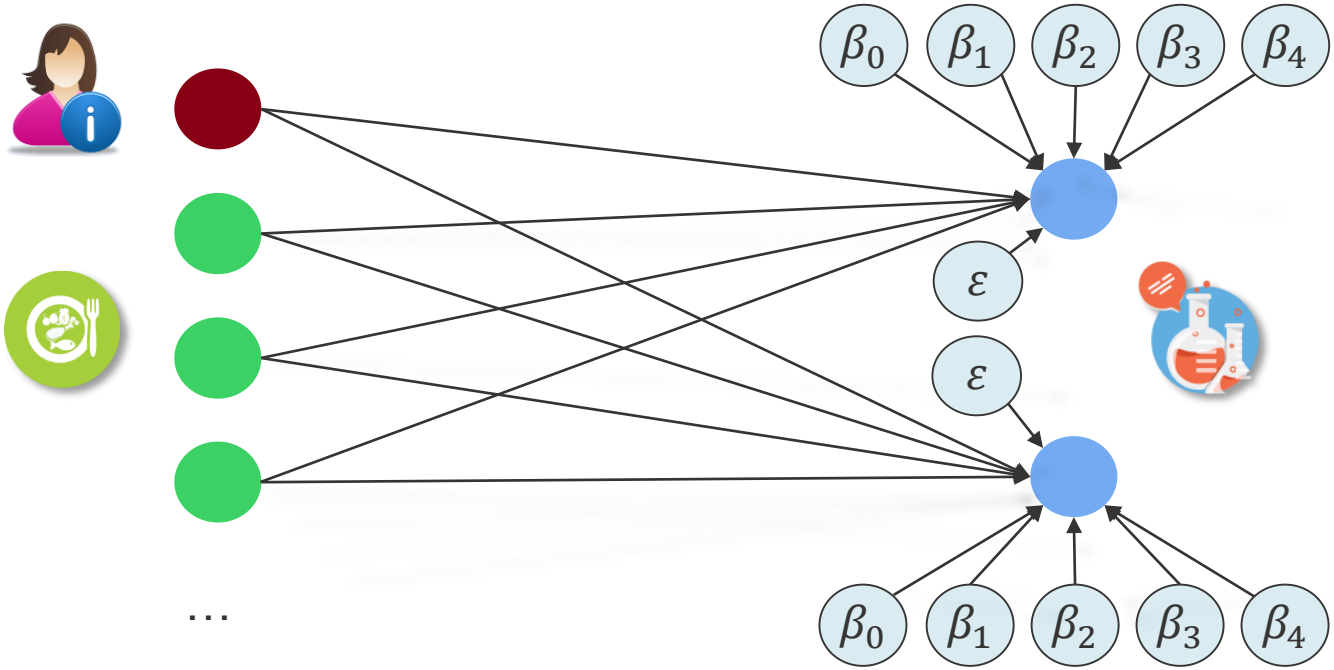
Towards personal understanding



So, it's a graph! Which graph describes the reactions best?

The graph extends to a graphical model

- The observed nodes are random variables from the exponential family
- Latent nodes indicate the strength and quality of the effect
- Linear link function is assumed between individual nutrients and response
- Optimal graph is searched amongst the all graphs, as $P(G|D) \propto P(D|G)P(G)$



- Prior information can be used to guide this graph search…

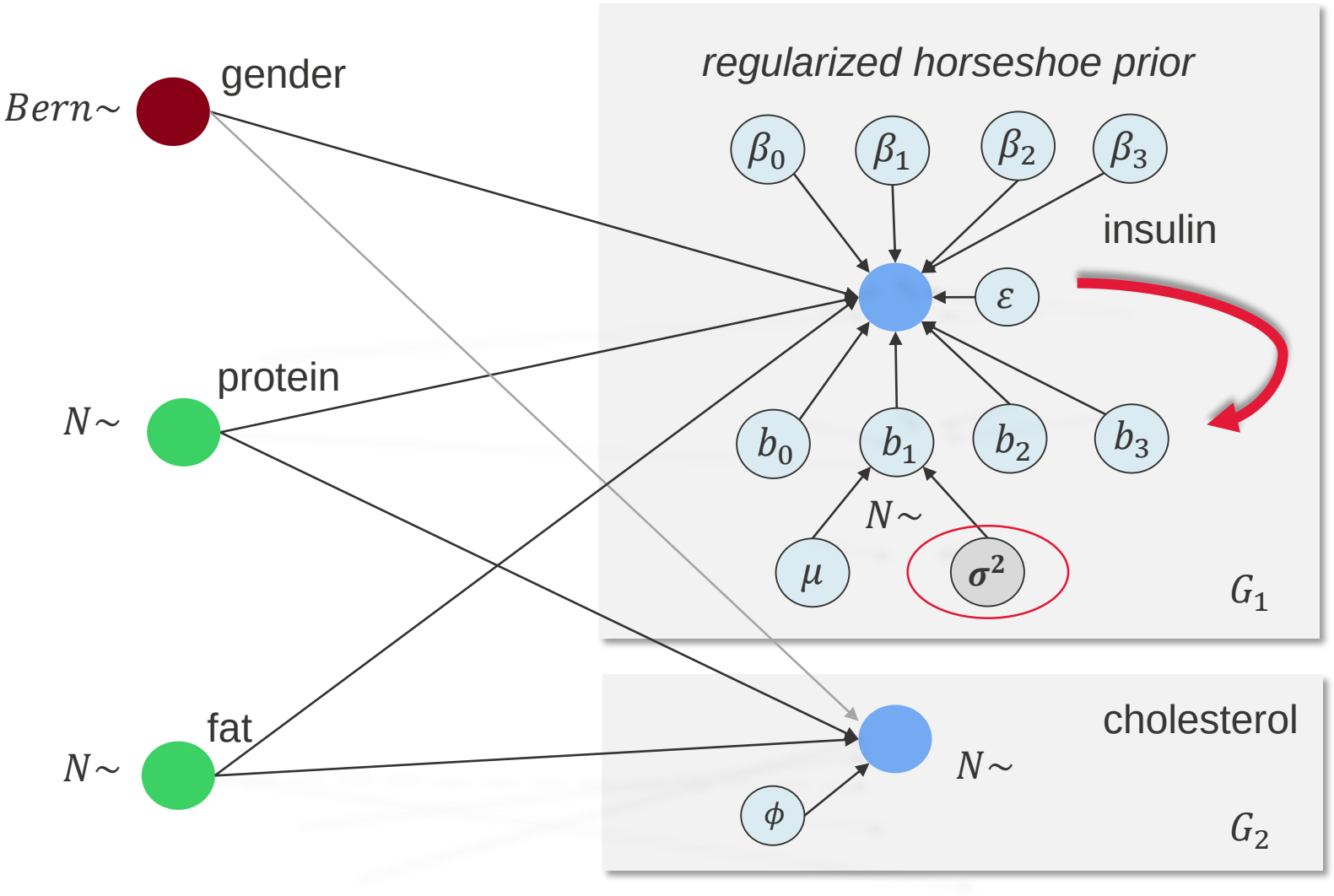
Prior information about the graphs – $P(G)$

- **Only biologically plausible graphs as considered, with some fixed $\boldsymbol{\beta = 0}$**
  - Assumption that nutrients are predictors of bodily responses
  - Nutrients don't affect each other, and responses don't affect nutrients
  - There might be some indirect connections, but BNs model these

- **Typical connections: structure learning with shrinkage**
  - We start from fully connected graph and prune out the edges with $\beta \approx 0$.
  - Regularized horseshoe prior (RHS) is used to enforce this shrinkage.
  - RHS allows specifying the number of effective predictors, i.e. nutrients that affect the response
  - Stricter prior graph could be also specified with prior values for betas

- **Typical nutrition and blood test levels**
  - Nutrients and responses can also have prior distributions
  - Informative prior could be taken from the current general guidelines that are actually specified with upper and lower bounds from Normal distribution

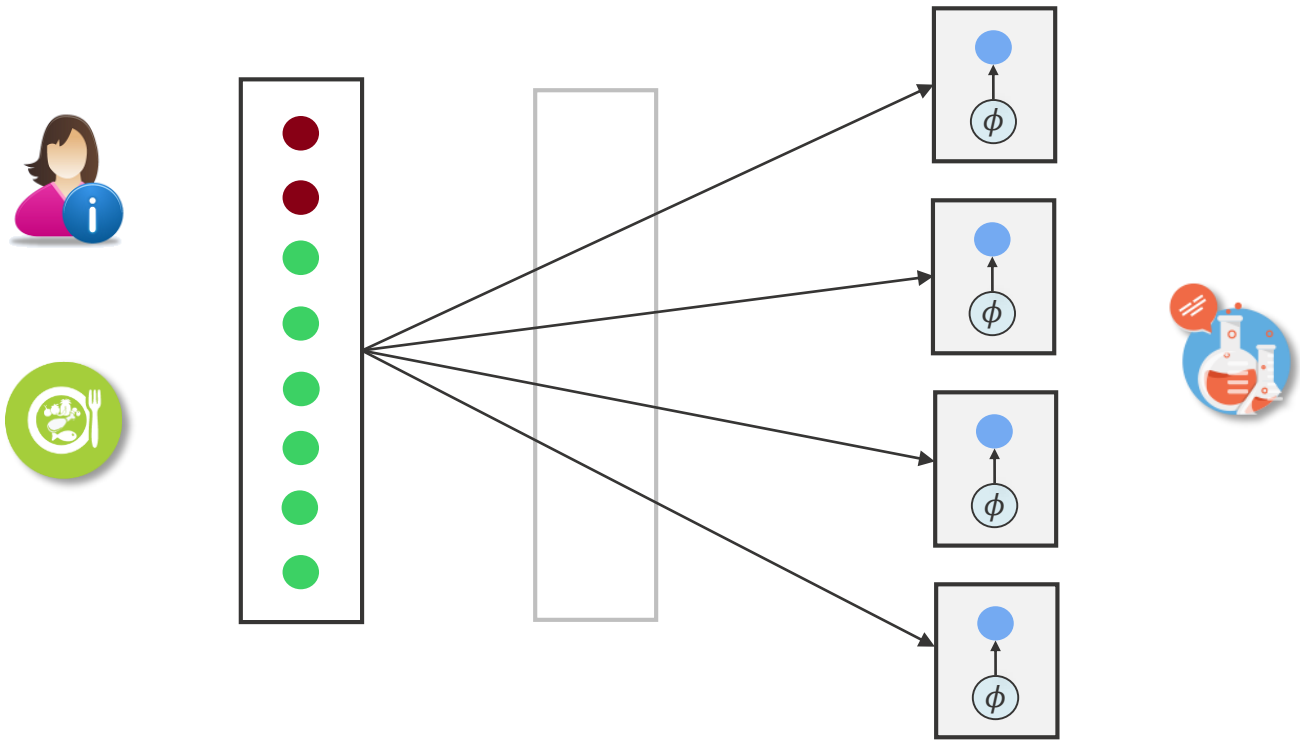
Adding hierarchy to local distributions - $P(D|G_i)$



Joint likelihood of the graph - $P(D|G)$

- Previous local distributions are evaluated for every non-root node
- Local distributions are assumed to be independent up to their Markov boundaries
- Joint probability of the graph is acquired by multiplying over all local distributions

$$P(D|G) = \prod P(X_i|pa(X_i), G_i)P(G_i)$$



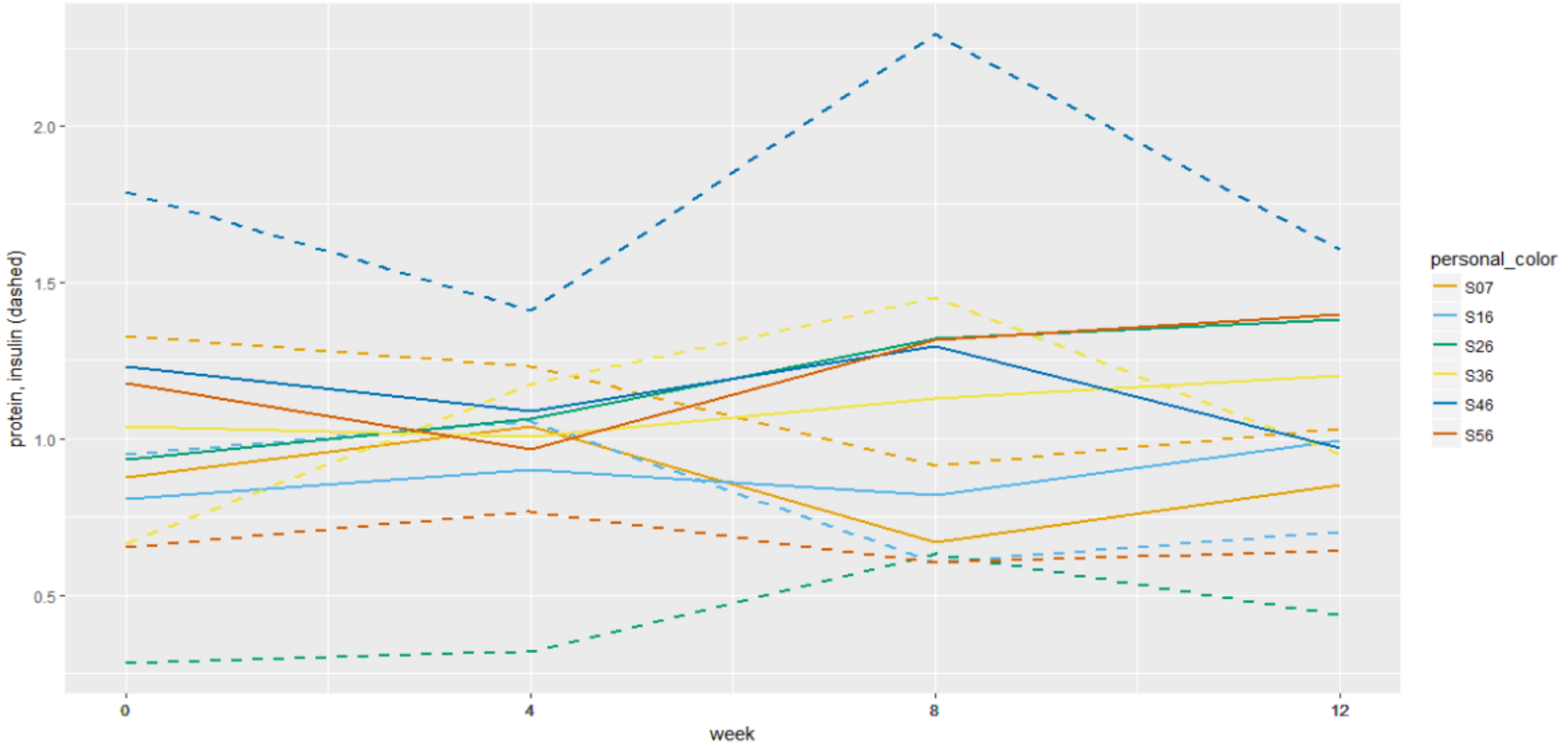
The data: Sysdimet study

- The original study is a randomized controlled trial with 106 persons in 3 groups and 1 control group for studying effects of the Nordic diet
- Besides the cholesterol medication there was no other prior knowledge about differences in personal reaction types to nutrition
- For this model we picked 17 nutrient variables, 5 blood test results, and medication and gender for personal information

- We have four observations during a 12 week period from food diary, blood tests and personal information
- Blood tests were taken always week after the food diary observation

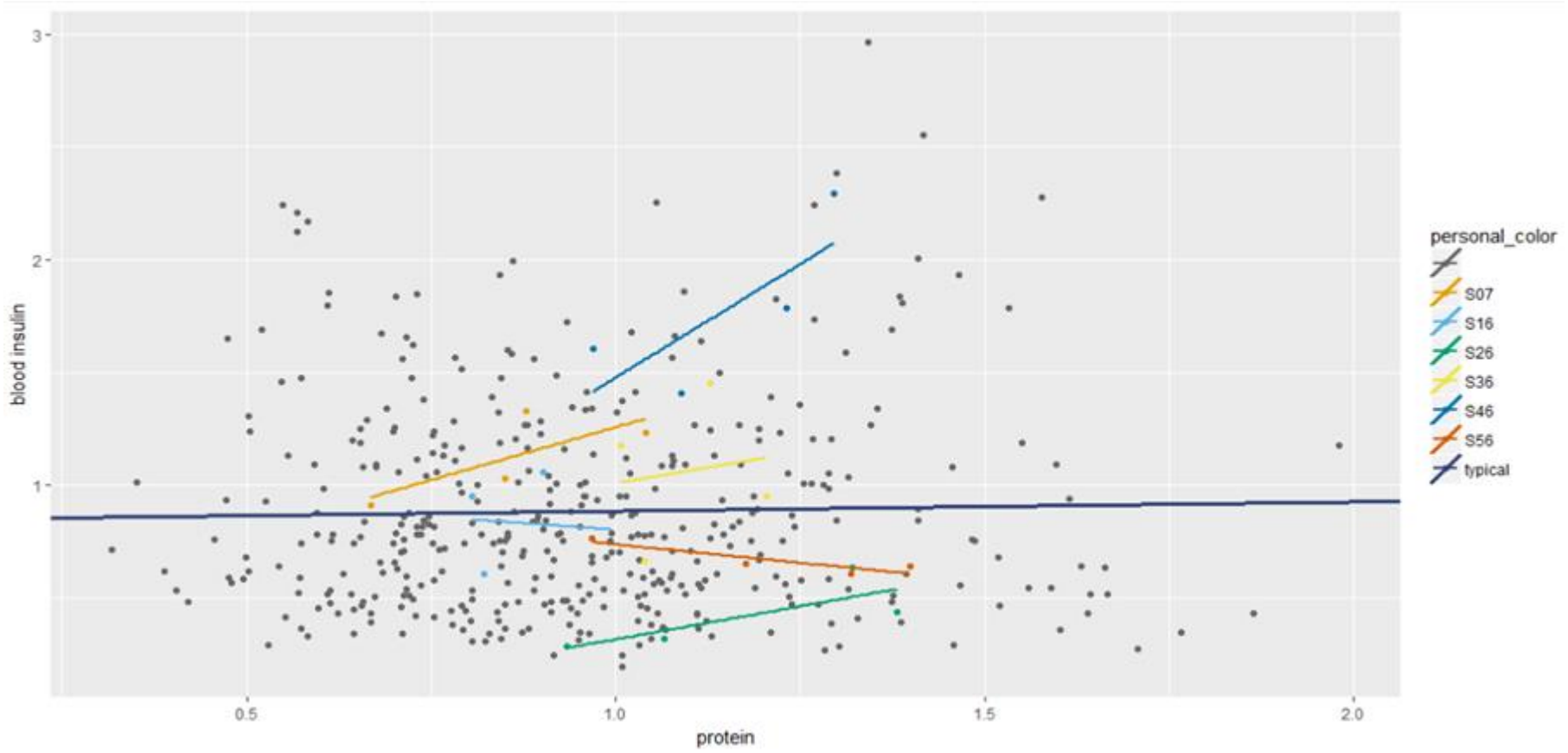
- Let's look at a small sample from data..

Data in Timeline: Protein's effect to insulin



- Let's take a running example of one nutritional effect: how the level of protein at diet affects to the person's blood insulin level.
- For longer timeseries, some autocorrelation structure, like moving average, could be added to the model.

Data in Scatter Plot: Protein's effect to insulin

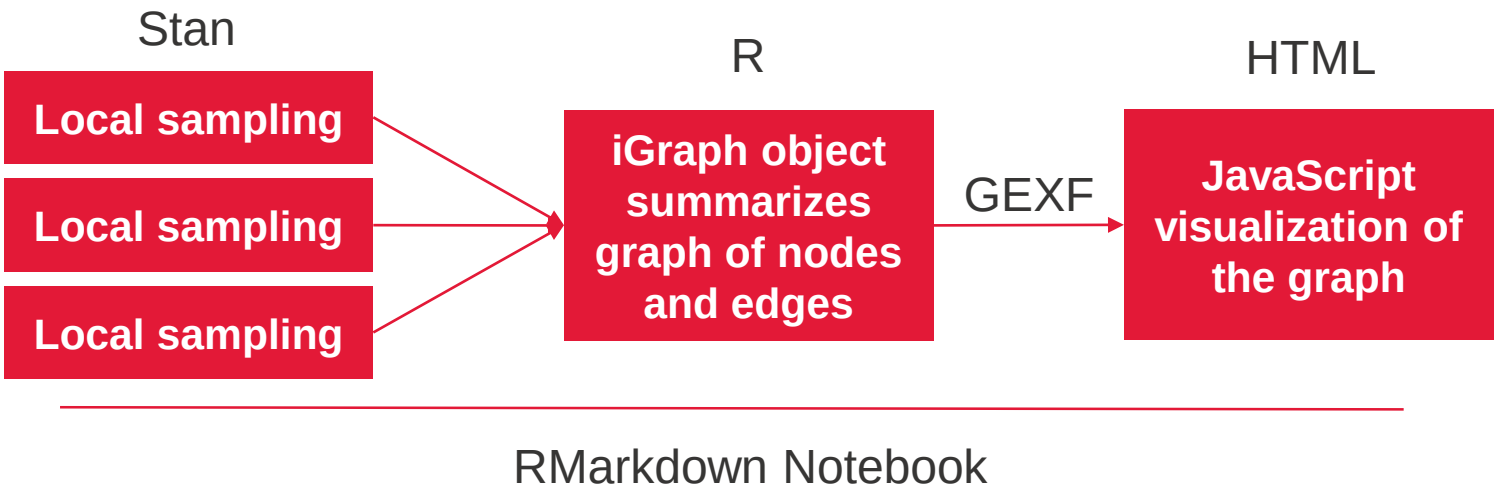


- These observations can be studied by fitting simple linear models to individual and all measurements. This resembles to our first graphical model with β only.
- We still consider the linear model adequate, but some of this noise can be explained with the personal effects.

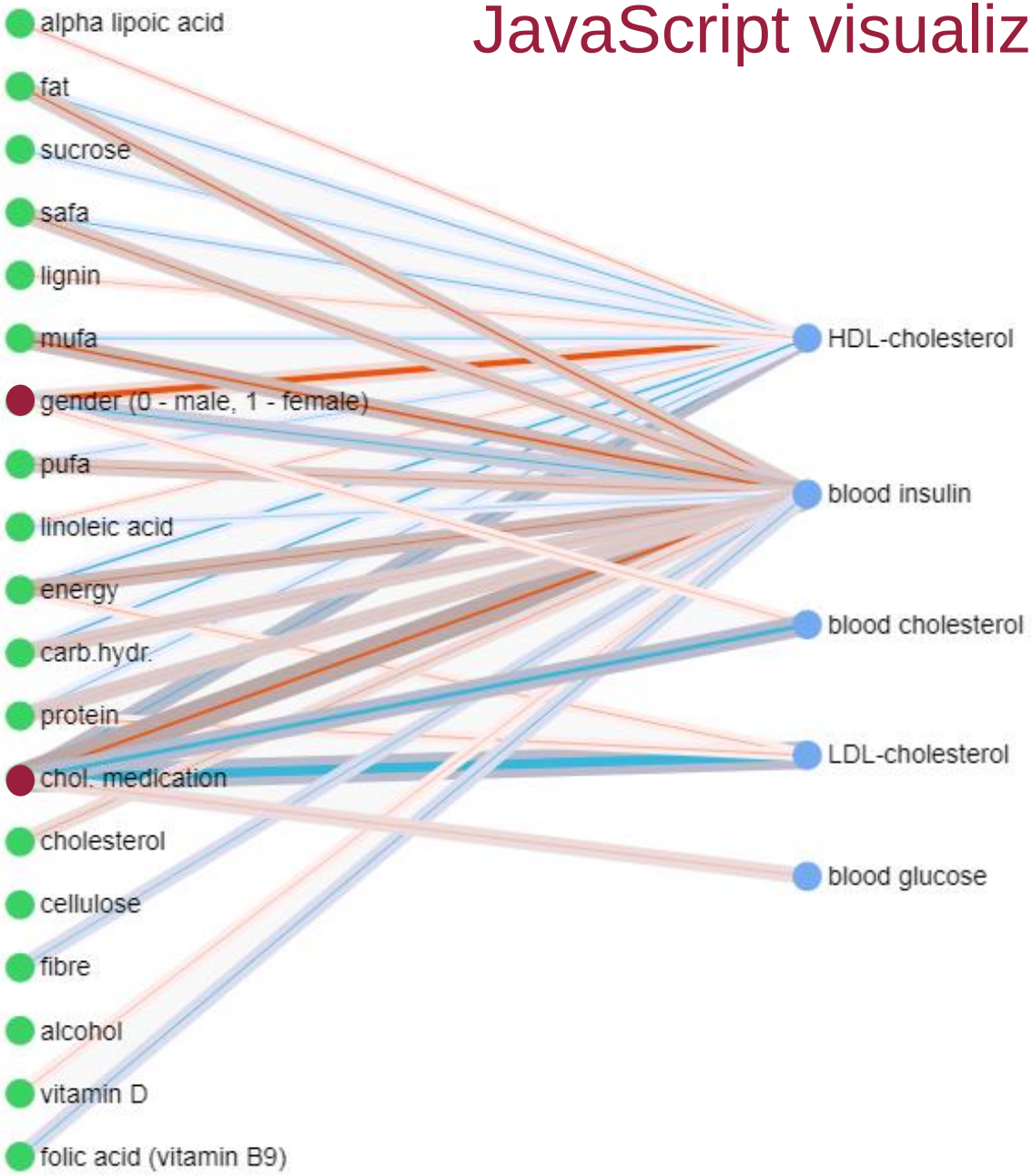
Implementation: Estimation with R and Stan

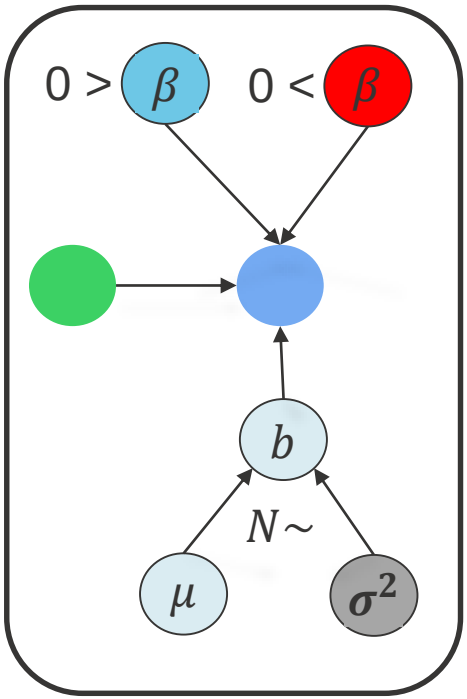
- The graph iteration is implemented with R code, and local distributions for every non-root node are estimated with Stan.
- The R code gathers estimation summaries and builds a graph in an iGraph object.

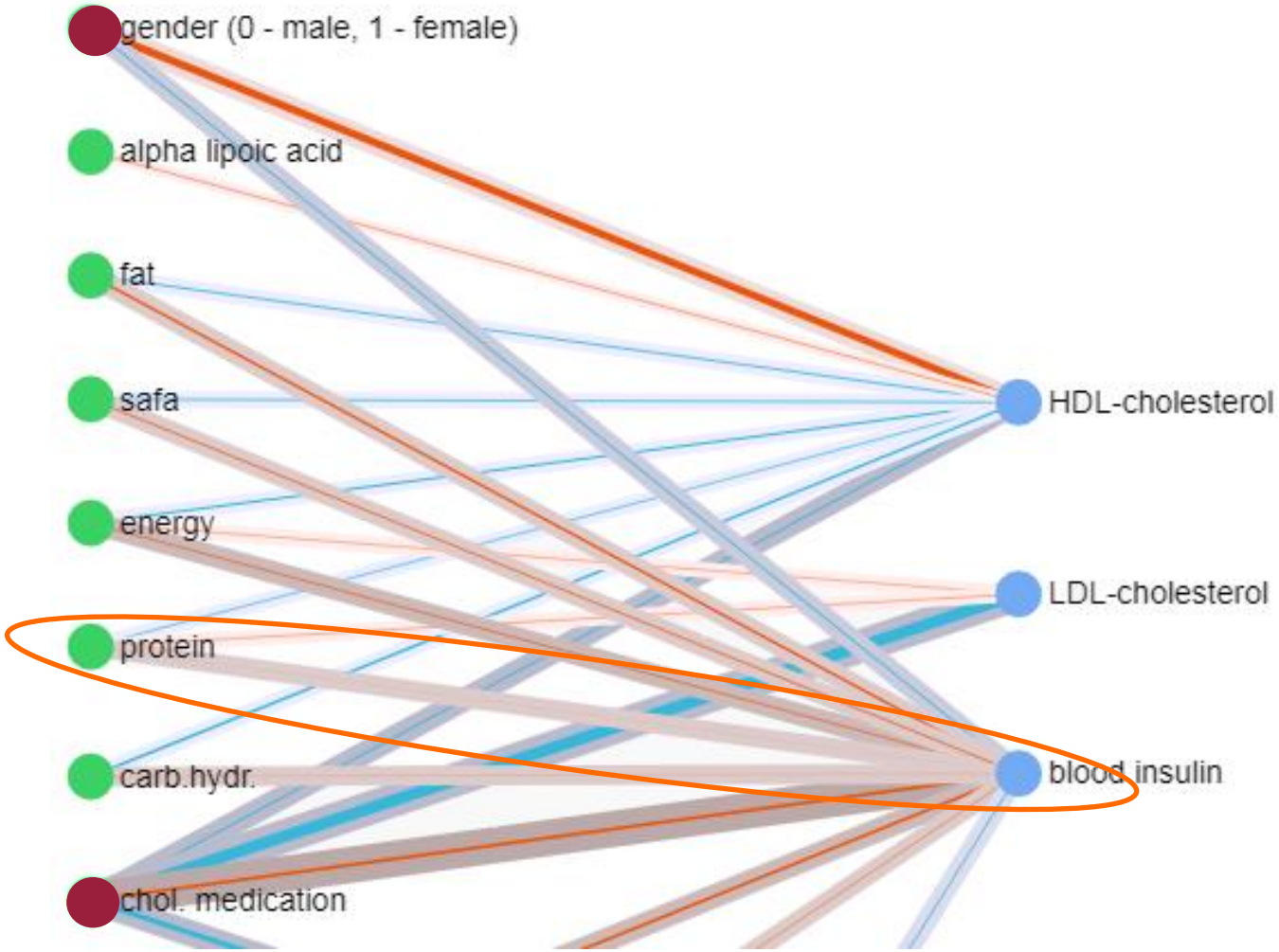
- For the Sysdimet dataset, this means that we have a graph with 19 nutrient or information nodes and 5 response nodes from blood tests.
- For regularizing horseshoe prior of beta, we are guessing that one third of the nutrients are relevant for each blood test result.

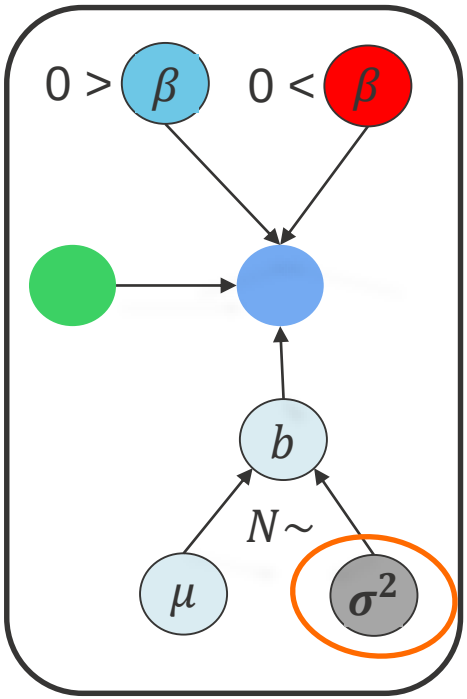


JavaScript visualization






gender (0 - male, 1 - female)
alpha lipoic acid
fat
safa
energy
protein
carb.hydr.
chol. medication
HDL-cholesterol
LDL-cholesterol
blood insulin


0 > β
0 < β
b
N ~
μ
σ²

Let's query the graph – for typical effects

From	To	Beta (normalized)
cholesterol medication	LDL-cholesterol	-0.026897912
cholesterol medication	blood cholesterol	-0.011823168
carbon hydrate	HDL-cholesterol	-0.006123904

From	To	Beta (normalized)
gender (female)	HDL-cholesterol	0.017239135
cholesterol medication	blood insulin	0.008108980
multisaturated fats (mufa)	blood insulin	0.007027948

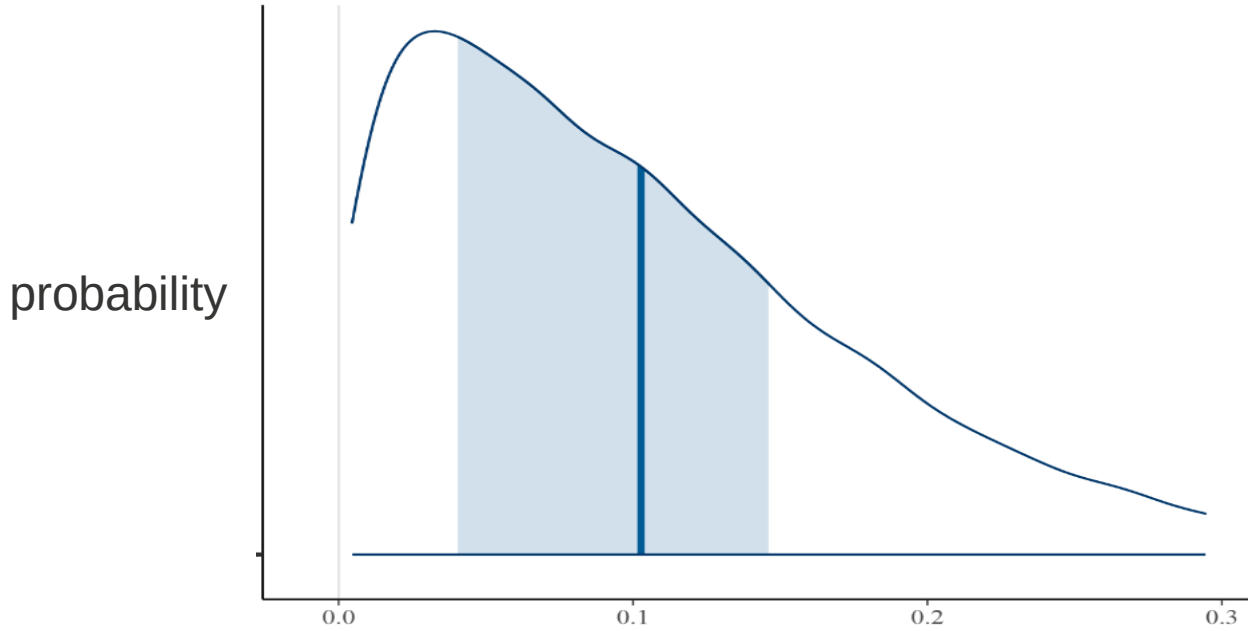
```
allnodes <- V(sysdimet_graph)
beta <- allnodes[allnodes$type=="beta"]
largest_typical_negative <- beta[order(beta$value),]
largest_typical_positive <- beta[order(-beta$value),]
```

… And for variances between persons

From	To	b_{σ^2}
energy	blood insulin	0.15997602
energy	LDL-cholesterol	0.10701078
energy	HDL-cholesterol	0.10680727
protein	**blood insulin**	**0.10266477**
pufa	blood insulin	0.10041444

```
allnodes <- V(sysdimet_graph)
b_sigma <- allnodes[allnodes$type=="b_sigma"]
largest_personal_variance <-
   b_sigma[order(-b_sigma$variance),]
```

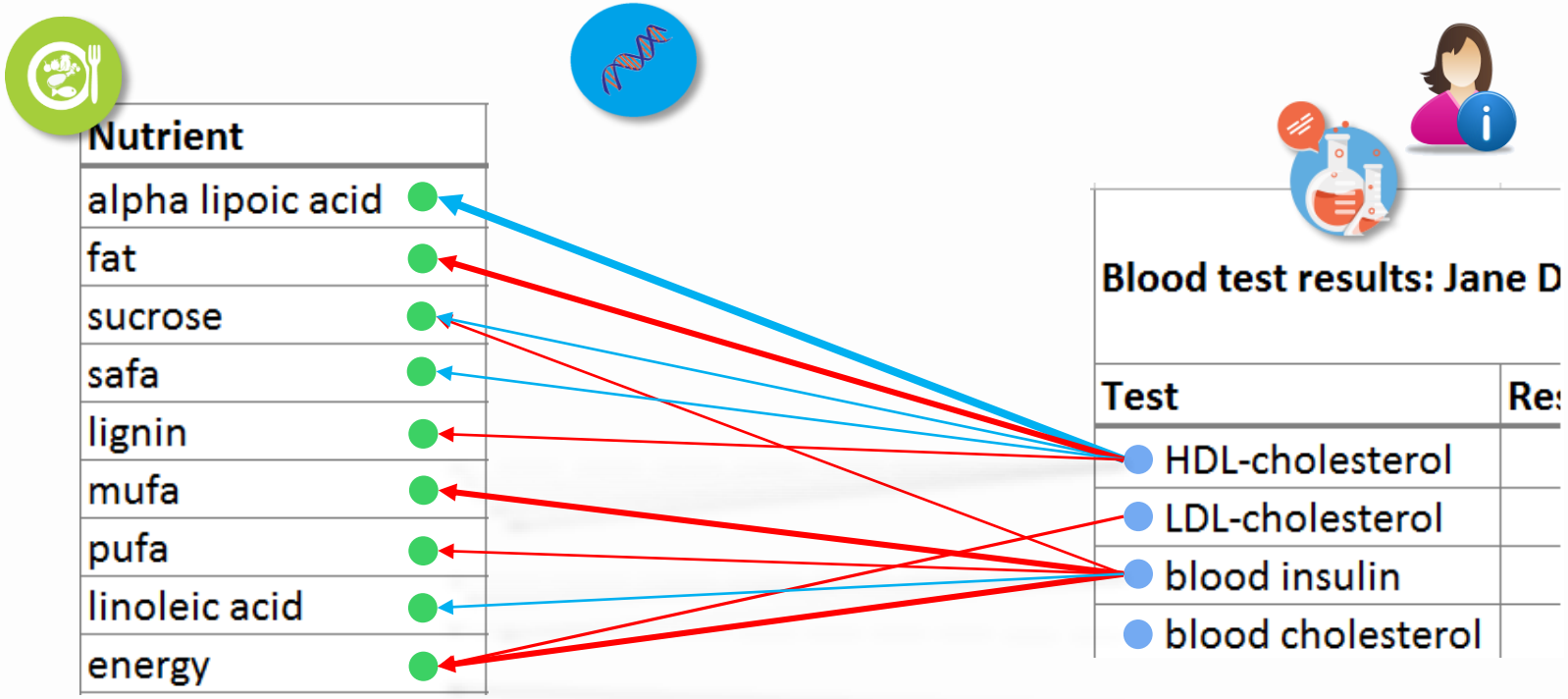
Querying the local distribution from Stan model: Personal variance in protein's effect to insulin



b_{σ^2}= 0.10266477

```
library("bayesplot")
fsins_posterior <- as.array(mebn.get_localfit("fsins"))
id <- match("prot", datadesc$Name)
mcmc_intervals(fsins_posterior, pars =
   c(paste0("sigma_b[",id,"]")), prob_outer = 0.95)..)
```

Future work: Personal graphs and diets



- With this model we can predict personal graphs for new patients
- Personal diet can be inferred by turning arrows at the personal graph and fixing the blood tests to their desirable levels
- Model can improved by adding the multi-response cases
- Studying new dataset with kidney disease patients and new variables

Thank you!

For more details or collaboration, contact me at jari.turkia@cgi.com