Center for Patient-Centered Outcomes Research

IBiKE

Correctly estimating prevalence with imperfect diagnostic tools

Felix Fischer & Dario Zocholl

Charité Universitätsmedizin Berlin

Learning objectives

- Understand the structure of a simple BLCMs for prevalence estimation
- Familiarize oneself with the peculiarities of prevalence estimation in depression
- Comprehend how the prior was generated from meta-analysis

- Fit BLCM in Stan
- Adjust the priors for the BLCM and compare results

- Understand that the data contains information about specificity
- Understand that diagnostic studies do not provide perfect/appropriate prior information

Prevalence of depression in Lancet Public Health

Prevalence and variability of current depressive disorder in 27 European countries: a population-based study

Jorge Arias-de la Torre, Gemma Vilagut, Amy Ronaldson, Antoni Serrano-Blanco, Vicente Martín, Michele Peters, Jose M Valderas, Alex Dregan, Jordi Alonso

Summary

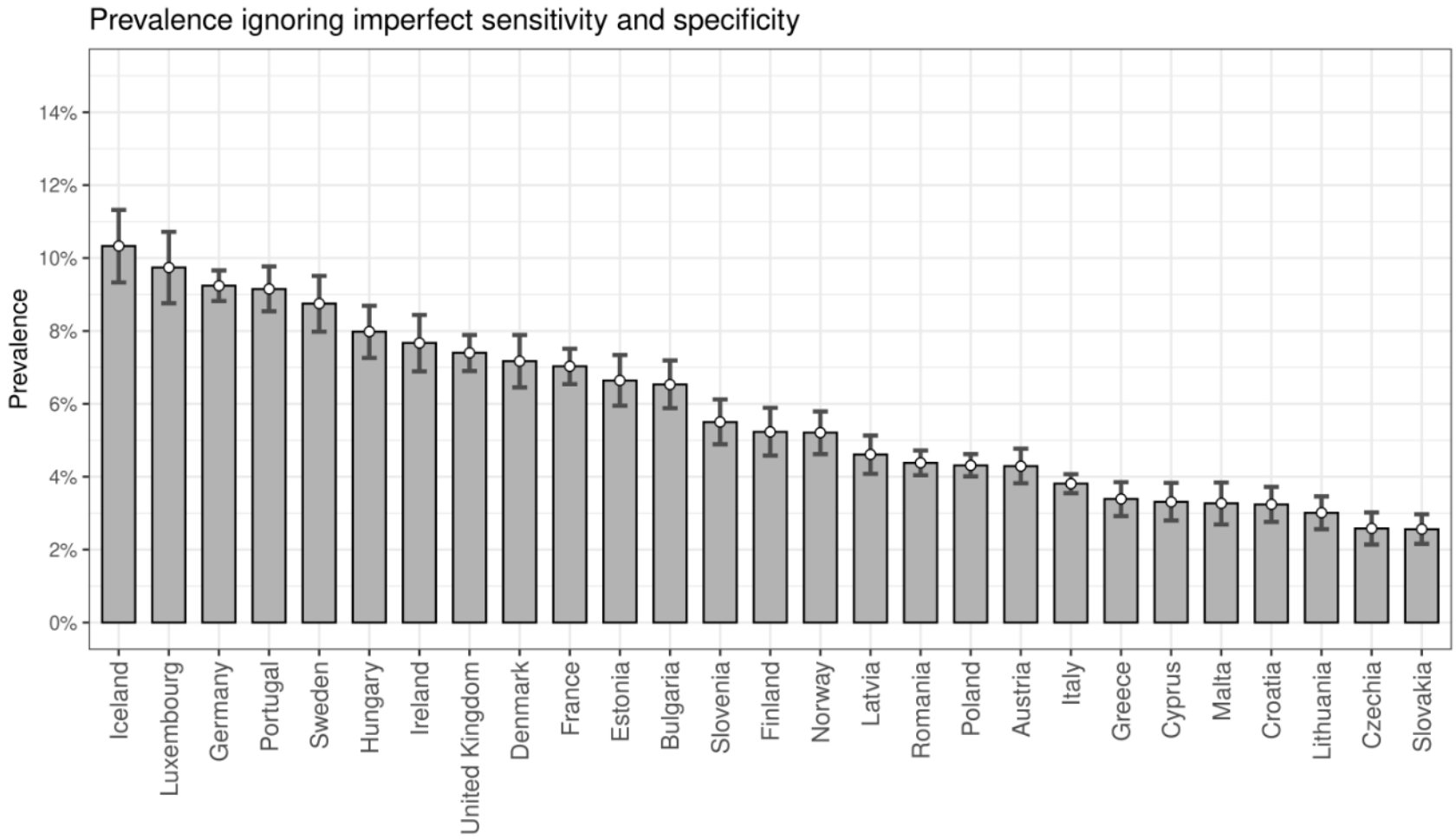
Background We aimed to estimate the prevalence of current depressive disorder in 27 European countries, and to explore differences in prevalence between European countries and by gender.

Methods In this population-based study, we analysed data from respondents living in 27 European countries who were included in the second wave of the European Health Interview Survey, collected between 2013 and 2015. We assessed the prevalence of current depressive disorder using the eight-item Patient Health Questionnaire (PHQ-8), with depressive disorder defined as a PHQ-8 score of 10 or higher. Prevalence estimates and 95% CIs were calculated for all 27 countries overall and for each country individually. We assessed variation in prevalence (country *vs* the rest of Europe) using crude and adjusted prevalence ratios obtained from negative binomial regression models. We did all analyses for the total sample and stratified by gender.

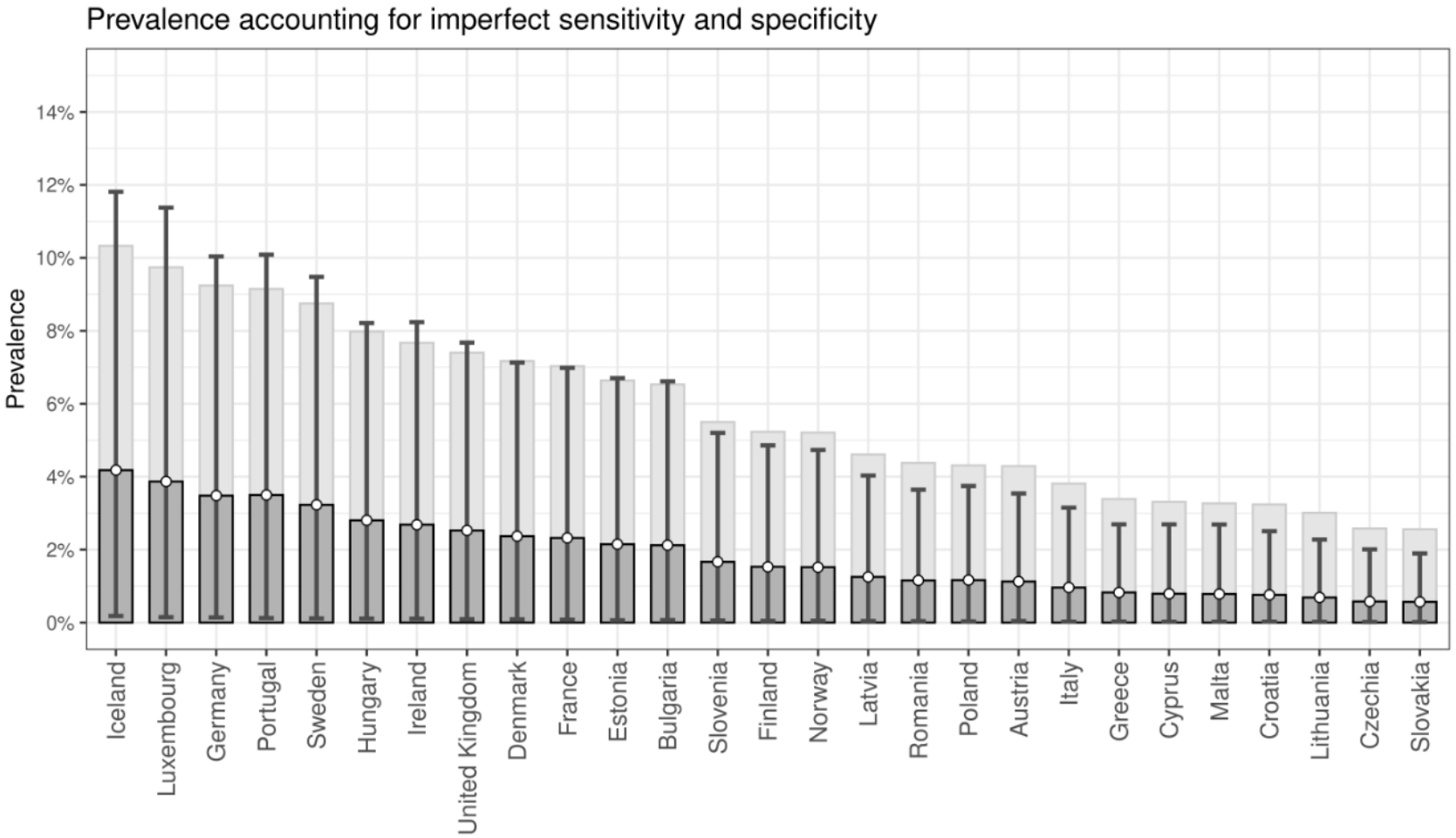
Findings Our analysis sample comprised 258 888 individuals, of whom 117 310 (weighted proportion 47·8%) were men and 141 578 (52·2%) were women. The overall prevalence of current depressive disorder was 6·38% (95% CI 6·24–6·52) with important variation across countries, ranging from 2·58% (2·14–3·02) in the Czech Republic to 10·33% (9·33–11·32) in Iceland. Prevalence was higher in women (7·74% [7·53–7·95]) than in men (4·89% [4·71–5·08]), with clear gender differences for all countries except Finland and Croatia. Compared with the other European countries in our sample, those with the highest adjusted prevalence ratios were Germany (1·80 [1·71–1·89]) and Luxembourg (1·50 [1·35–1·66]), and those with the lowest adjusted prevalence ratios were Slovakia (0·28 [0·24–0·33]) and the Czech Republic (0·32 [0·27–0·38]).

Interpretation Depressive disorders, although common across Europe, vary substantially in prevalence between countries. These results could be a baseline for monitoring the prevalence of current depressive disorder both at a country level in Europe and for planning health-care resources and services.

Ignoring imperfect accuracy



Considering imperfect accuracy



PHQ-8 as diagnostic tool is not perfect

Psychological Medicine

cambridge.org/psm

Original Article

Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis‡

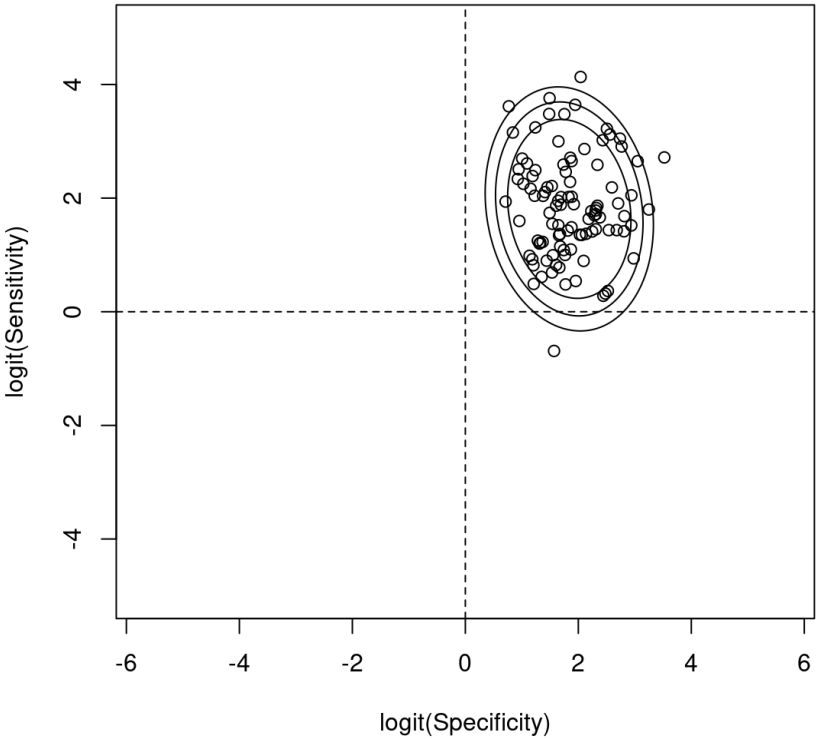
Yin Wu[1–3], Brooke Levis[1,3], Kira E. Riehm[1], Nazanin Saadat[1], Alexander W. Levis[1], Marleine Azar[1], Danielle B. Rice[1,4], Jill Boruff[5], Pim Cuijpers[6], Simon Gilbody[7],

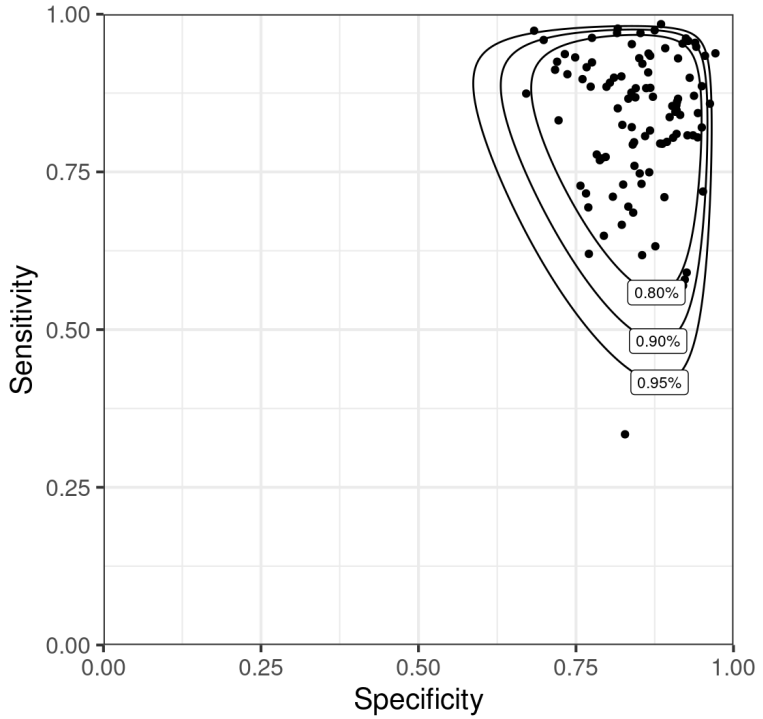
- 27 included primary studies that used semi- structured interviews to assess major depression (6362 participants/790 cases)
- Cutoff 10 maximized combined sensitivity and specificity for PHQ-8 [sensitivity (95% CI) = 0.86 (0.80–0.90), specificity (95% CI) = 0.86 (0.83–0.89)]
- Bivariate generalized random-effects model (Riley et al. 2008) to jointly estimate sensitivity and specificity
- Bivariate normal distribution of sensitivity and specificity on the logit scale

Sensitivity and specificity

A multivariate normal prior distribution of sensitivity and specificity can be specified based on the estimated mean logit-sensitivity, mean logit-specificity, between-study variances and between-study correlation from this bivariate generalized logistic linear model.

$$\begin{pmatrix} logit\,(Se_i) \\ logit\,(Sp_i) \end{pmatrix} \sim N\left[\begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \tau_1\tau_0\rho \\ \tau_1\tau_0\rho & \tau_0^2 \end{pmatrix}\right] \qquad N\left[\begin{pmatrix} 1.84 \\ 1.81 \end{pmatrix}, \begin{pmatrix} 0.37 & -0.07 \\ -0.07 & 0.77 \end{pmatrix}\right]$$





Introduction to estimating prevalences

- Given the prevalence, the sensitivity and the specificity, the positive rate p can be calculated as:

$$p = prev * Se + (1 - prev) * (1 - Sp)$$

- Alternatively, solving for the prevalence, we can write:

$$prev = (p + Sp - 1)/(Se + Sp)$$

- So, we can easily calculate the prevalence from the observed positive rate $\hat{p}$, the sensitivity and the specificity.

- Some difficulties arise:
  1. If $\hat{p} < 1 - Sp$, the expression for $prev$ becomes meaninglessly negative.
  2. If Se and Sp are not known, the equation is unsolable.
  3. Incorporating uncertainty of Se and Sp is difficult due to the non-linear relation.

Introduction to estimating prevalences

- Given the prevalence, the sensitivity and the specificity, the positive rate p can be calculated as:

$$p = prev * Se + (1 - prev) * (1 - Sp)$$

- Alternatively, solving for the prevalence, we can write:

$$prev = (p + Sp - 1)/(Se + Sp - 1)$$

- So, we can easily calculate the prevalence from the observed positive rate $\hat{p}$, the sensitivity and the specificity.

- Some difficulties arise:
  1. If $\hat{p} < 1 - Sp$, the expression for $prev$ becomes meaninglessly negative.
  2. If Se and Sp are not known, the equation is unsolvable.
  3. Incorporating uncertainty of Se and Sp is difficult due to the non-linear relation.

  → We will need Bayesian statistics here.

  Teaser: when done correctly, the estimate of depression prevalence e.g. in Iceland is not 10.3% but approx. 4%, and for all countries uncertainty is so high that it is basically something between 0% and 12%.

Bayesian model to estimate prevalence

Assuming J countries, the number of positive tests in each country y_j with j $\in \{1, \dots, J\}$ follows a binomial distribution with tested individuals n_j and positive test rate p_j:

$$y_j \sim Binomial(n_j, p_j)$$

Assuming priors for sensitivitiy and specificity:

$$Sens_j \sim Beta(a, b)$$
$$Spec_j \sim Beta(a, b)$$

as well as a prior for the prevalence:

$$prev_j \sim Beta(a, b)$$

the posterior distribution of the prevalence can be calculated.

Specification of priors

- A recently individual participant data meta-analysis estimated the parameters of the multivariate lognormal distribution of sensitivity and specificity:

$$\begin{pmatrix} logit\ (Se_j) \\ logit\ (Sp_j) \end{pmatrix} \sim N\left[\begin{pmatrix} 1 \cdot 84 \\ 1 \cdot 81 \end{pmatrix}, \begin{pmatrix} 0 \cdot 37 & -0 \cdot 07 \\ -0 \cdot 07 & 0 \cdot 77 \end{pmatrix}\right]$$

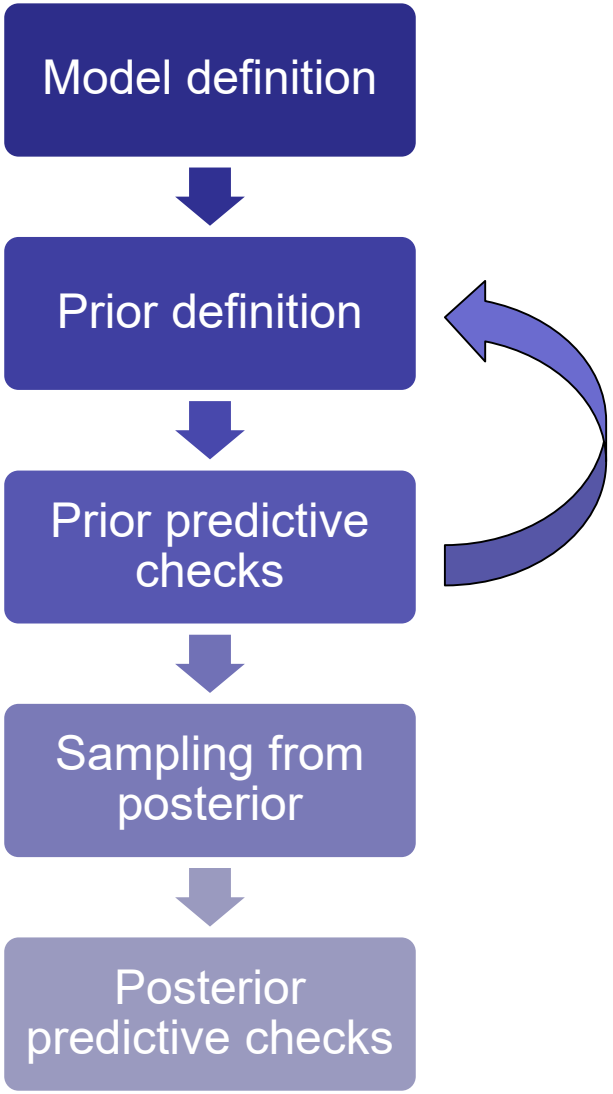
- This corresponds to 86% specificity and 84% sensitivity.
- We use this as starting point for our prior.

- Additionally, for the prevalence we do not make any informative assumption, so we place a uniform distribution on it:

$$Prev_j \sim Beta(1,1)$$

- Now, we have everything we need to specify the model.

Principle Bayesian Workflow



Model definition

- The model consists of:
  - the formula for calculation of a positive test rate:
$$p = prev * Se + (1 - prev) * (1 - Sp)$$
  - the central sampling statement:
$$y \sim Binomial(n, p)$$
  - a definition of the model parameters, i.e. a statement that the following parameters are to be modelled and therefore are assumed to have a prior distribution:
    - Se
    - Sp
    - prev
    - Note: p is not a model parameter because it is observed data

Definition of priors

- The priors for sensitivity and specificity are taken from the meta-analysis as described previously.
- As prior for the prevalence, a uniform prior might be applied to be weakly informative.
- However, we should remember that we are modeling the positive test rate, which is a function of all three parameters,
  therefore we must not think of the priors as being separate entities, but must consider them always being interdependent.
- This is even more important, since in our data there is no information about sensitivity and specificity at all.
- One way to get an understanding of this interdependency are prior predictive checks.

Prior predictive check

- A prior should always represent a belief about the data that appears meaningful.
- A prior predictive check generates data from the prior in order to assess whether it is a useful model of our a priori beliefs about the data.
- This becomes particularly relevant when there are parameters which result from a combination of several priors.
  - e.g. sensitivity, specificity and prevalence, which together determine the positive test rate.
- Important: the actually observed data must not play any role in the prior predictive check!
- If the generated data does not appear to be realistic in any sense, it may be necessary to adjust the priors.

Prior predictive check - example

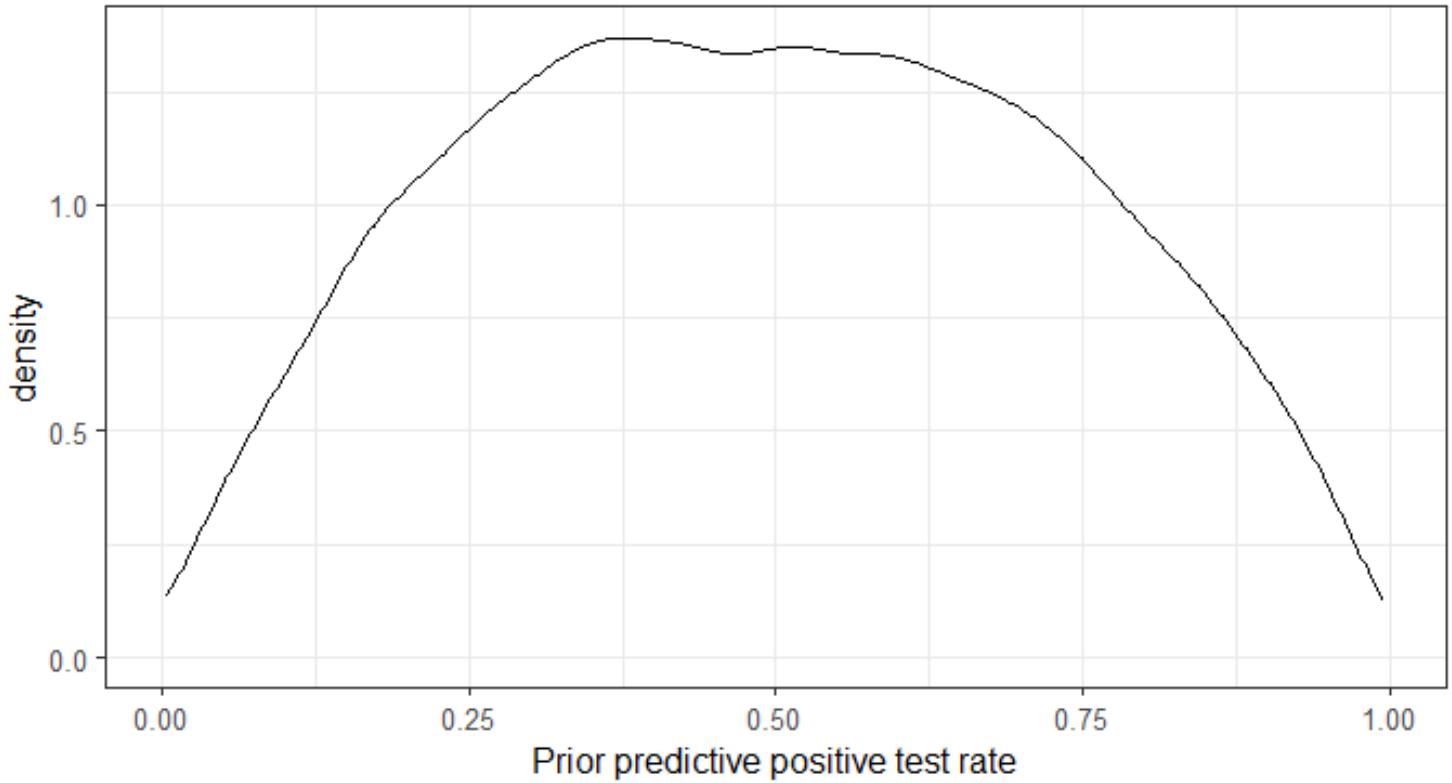
- Take the „default non-informative“ prior Beta(1,1) for each parameter, i.e.: $prev \sim Beta(1,1)$, $\mathrm{Se} \sim Beta(1,1)$, $Sp \sim Beta(1,1)$
- A prior predictive check can easily be done in R:
  - Sampling from the prior distributions:

```
prev <- rbeta(n=1e4, shape1=1, shape2=1)
Se   <- rbeta(n=1e4, shape1=1, shape2=1)
Sp   <- rbeta(n=1e4, shape1=1, shape2=1)
```

  - Transforming the generated parameter samples to get samples from the prior predicted data:

```
pos_test_rate <- prev * Se + (1-prev) * Sp
```

Prior predictive check - example



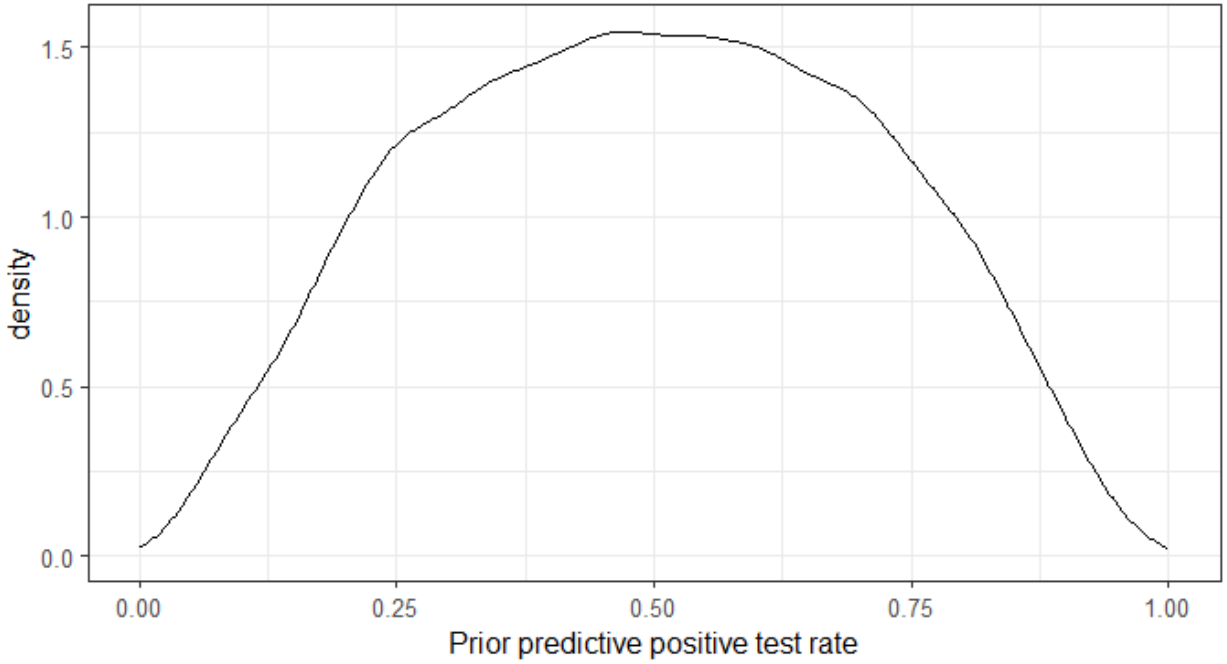
- Let‘s think about it for a second: if we did not have any knowledge about the actual data, would we think the most likely values for the positive test rate would be around 50%?

Prior predictive check - example

- We can use some slightly more informative priors.
- Sensitivity and specificity might be expected to be around 90%, similar the observation in the meta-analysis previously mentioned.

```
Se   <- rbeta(n=1e4, shape1=10, shape2=2)
Sp   <- rbeta(n=1e4, shape1=10, shape2=2)
```

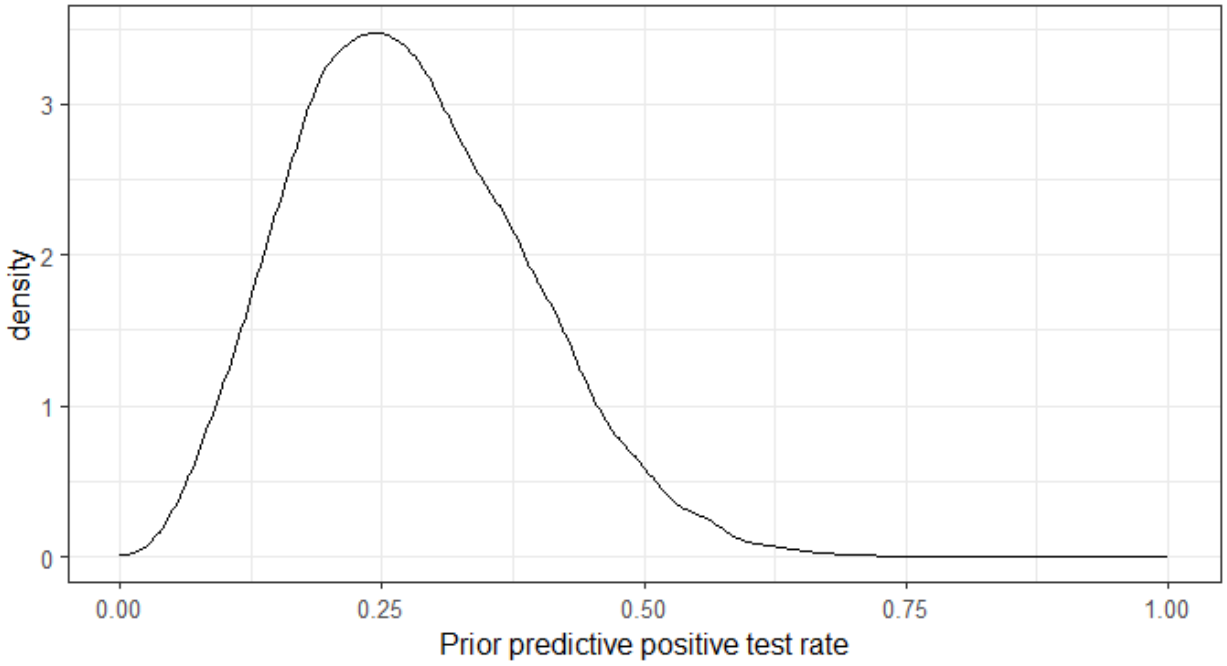
- However, this does not actually change much:



Prior predictive check - example

- For the prior predictive distribution to be meaningful, we actually need to consider being (seemingly) more informative for the prevalence.
- For example:

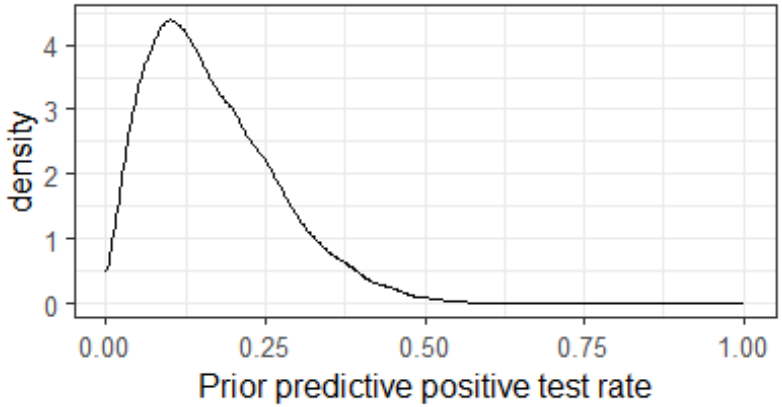
```
prev <- rbeta(n=1e4, shape1=2, shape2=10)
Se   <- rbeta(n=1e4, shape1=10, shape2=2)
Sp   <- rbeta(n=1e4, shape1=10, shape2=2)
```



Prior predictive check - example

- Assume for a moment, we want to express a huge amount of information in our prevalence prior, for example:

```
prev <- rbeta(n=1e4, shape1=2, shape2=1000)
Se   <- rbeta(n=1e4, shape1=10, shape2=2)
Sp   <- rbeta(n=1e4, shape1=10, shape2=2)
```



- Note that this prior corresponds to only 1 truly pos case in 1000 samples.
- The prior predictive positive test rate is still quite high.
  How is that possible?

Posterior predictive checks

- The logic of posterior predictive checks is anaogously to prior predictive checks.
- The question to be answered is: does the model fit to the data?
- A posterior predictive distribution is the distribution of outcomes (i.e. positive test rates) based on the posterior distribution of the parameters and the observed data (i.e. n).
- It should be consistent with the actual observations.
- Otherwise, this may indicate that model assumptions have actually not been fulfilled by the data, or that the priors heavily influence the posterior inference.

Hands-on session

1. Run the prior predictive check as shown in the slides.
2. Assume that we have a strong prior belief that the positive test rates are below 5%. How can this be achieved?
3. Perform a prior predictive check with the bivariate normal prior:

$$\begin{pmatrix} logit\,(Se) \\ logit\,(Sp) \end{pmatrix} \sim N\left[\begin{pmatrix} 1 \cdot 84 \\ 1 \cdot 81 \end{pmatrix}, \begin{pmatrix} 0 \cdot 37 & -0 \cdot 07 \\ -0 \cdot 07 & 0 \cdot 77 \end{pmatrix}\right]$$

Hint: samples from a multivariate normal distribution can be generated by using the function `rmvnorm()` from the `mvtnorm`-package, and these have to undergo an inverse logit transformation to transform from [–Inf, Inf] to [0,1], e.g.

```
inv_logit <- function(x) exp(x)/(1+exp(x))
Se <- inv_logit(Se_Sp[,1])
```

   a. Start with a uniform Beta prior for the prevalence. Are the results reasonable?
   b. Adjust the prior for the prevalence to see whether there may be more reasonable specifications.
   c. Based on your (subjective) judgement, is a modification of the priors for sensitivity and specificity required?

Stan

Stan - Introduction

- Stan is a probabilistic modeling language written in C++ dedicated for conducting Bayesian modelling and inference, loosely based on BUGS
- Uses highly efficient Hamiltonian Monte Carlo and No-U-Turn sampler (NUTS)
- Created by Andrew Gelman and Bob Carpenter + development team
- Interfaces available in R, Python, MATLAB, Julia, Stata, Mathematica, and for the command line
- Interface to R via package rstan
- Github repository: github.com/stan-dev/stan
- Elaborate User's guide: https://mc-stan.org/docs/2_23/stan-users-guide/index.html#overview
- Highly responsive and active community: https://discourse.mc-stan.org/

Stan – code blocks (basic)

```
data
  // data passed by the user
parameters
  // unkown variables to be sampled
transformed parameters
  // variables depending on data and parameter blocks
model
  // specification of the log-posterior density
```

Stan – code blocks

functions
  // user defined Stan functions
data
  // data passed by the user
transformed data
  // variables depending on the data block
  // computed only once before fitting the model
parameters
  // unkown variables to be sampled
transformed parameters
  // variables depending on data and parameter blocks
  // computed for each iteration of posterior draws
model
  // specification of the log-posterior density
generated quantities
  // variables to be computed after the model fitting
  // not included in the actual sampling process

Stan – data types

```
// real numbers
real mu;
// integer
int y;
// vector of size n: list of real numbers
vector[n] v;
// matrix: table of real numbers
matrix[2,4]
// array: list of any type of data
int a[4];
//arrays can be of higher dimension
int a[2,4]

// boundaries on real numbers, [0, ∞]
real<lower=0> sigma;
// boundaries on real numbers, [0,22]
real<lower=0, upper=22> sigma;
```

Many specific data types, e.g. corr_matrix[3] sigma → 3x3 correlation matrix

Simple prevalence model

```
data {
  int<lower = 0> n; // number of tested persons
  int<lower = 0> y; // number of positive persons
  vector<lower = 0>[2] beta_prev; // parameters for beta distribution
  vector<lower = 0>[2] beta_sens; // parameters for beta distribution
  vector<lower = 0>[2] beta_spec; // parameters for beta distribution
}
parameters {
  real<lower = 0, upper = 1> prev;
  real<lower = 0, upper = 1> sens;
  real<lower = 0, upper = 1> spec;
}
transformed parameters {
  real<lower = 0, upper = 1> p;
  p = prev * sens + (1 - prev) * (1 - spec);
}
model {
  y ~ binomial(n, p);
  prev ~ beta(beta_prev[1],beta_prev[2]);
  sens ~ beta(beta_sens[1],beta_sens[2]);
  spec ~ beta(beta_spec[1],beta_spec[2]);
}
```

Prevalence model with bivariate normal prior

```
data {
  int<lower = 0> n; // number of tested persons
  int<lower = 0> y; // number of positive persons
  vector[2] mu;
  vector<lower = 0>[2] prev_beta;
  cov_matrix[2] Sigma;
}
parameters {
  real<lower = 0, upper = 1> prev;
  vector[2] logit_Se_Sp;
}
transformed parameters {
  real logit_Se = logit_Se_Sp[1];
  real logit_Sp = logit_Se_Sp[2];

  real Se = inv_logit(logit_Se);
  real Sp = inv_logit(logit_Sp);

  real<lower = 0, upper = 1> p;
  p = prev * Se + (1 - prev) * (1 - Sp);
}
model {
  y ~ binomial(n, p);
  logit_Se_Sp ~ multi_normal(mu, Sigma);
  prev ~ beta(prev_beta[1],prev_beta[2]);
}
```

Running Stan via R

In order to run the model, the following steps need to be taken:

1. Save the model as .stan file (supported by RStudio).
2. Compile the model (in R) via
   ```
   prevalence_model <- stan_model("prevalence_model_practise.stan")
   ```
   this might take 1-2 minutes.
3. Define the data to be passed to the model, e.g.:
   ```
   stan_data <- list(n=24404, y=2269, beta_prev=c(1,1),
   beta_sens=c(1,1), beta_spec=c(1,1))
   ```
   or for the model with bivariate normal prior:
   ```
   stan_data <- list(n=24404, y=2269, mu=c(1.84,1.81),
   Sigma=matrix(c(0.77,-0.07,-0.07,0.37), nrow=2), prev_beta=c(1,1))
   ```
4. Sample from the posterior distribution:
   ```
   fit <- sampling(prevalence_model, data=stan_data, chains=4,
   iter=2000, cores=1)
   ```
5. Pay attention to any warning messages. (check the .html file for additional information)

Hands-on session

- In the European Health Survey, the PHQ-8 was collected as a measure of depression severity. The following table shows the observed results for Germany:

Test positive (PHQ-8 >= 10)	Test negative (PHQ-8 < 10)	Total
2,269	22,135	24,404

1. Estimate the posterior prevalence of depression given this data using a uniform prior on prevalence (beta(1,1)) and informative priors on sensitivity (beta(87,15)) and specificity (beta(87, 15)).
2. Run the model with bivariate normal priors.
3. What do you notice about the posterior estimates of specificity?
4. Prior and posterior predictive checks can be done directly in Stan by adding a new block, e.g. at the end of the Stan file:

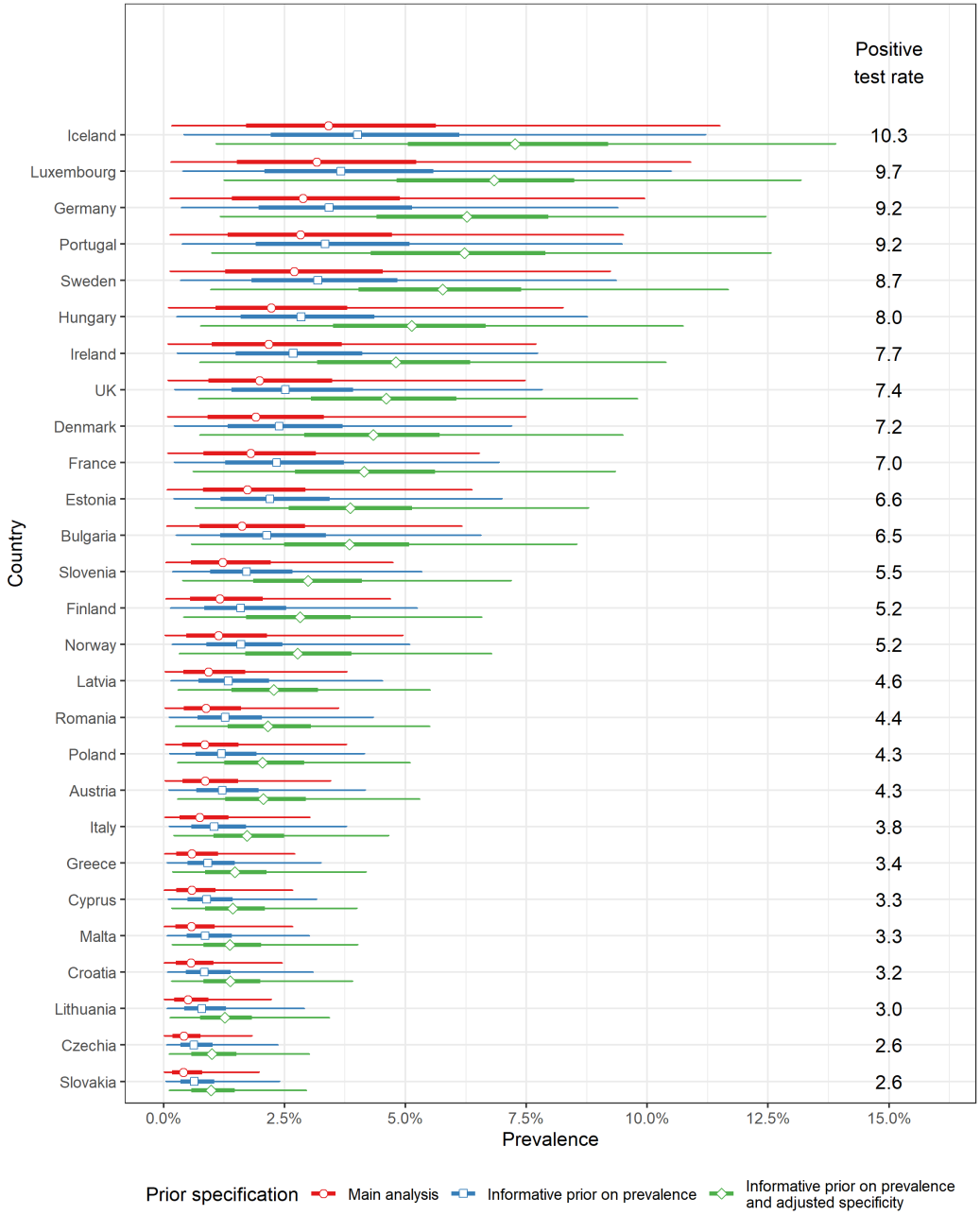
```
generated quantities {
  int y_pred = binomial_rng(n, p);
}
```

   Recall the prior predictive checks from earlier, can you perform them via Stan? Hint: “n” should still be passed to Stan, but not “p” (positive tests).
5. What impact do different priors have on the posterior distribution? Run a prior sensitivity analysis by comparing the results with different priors.

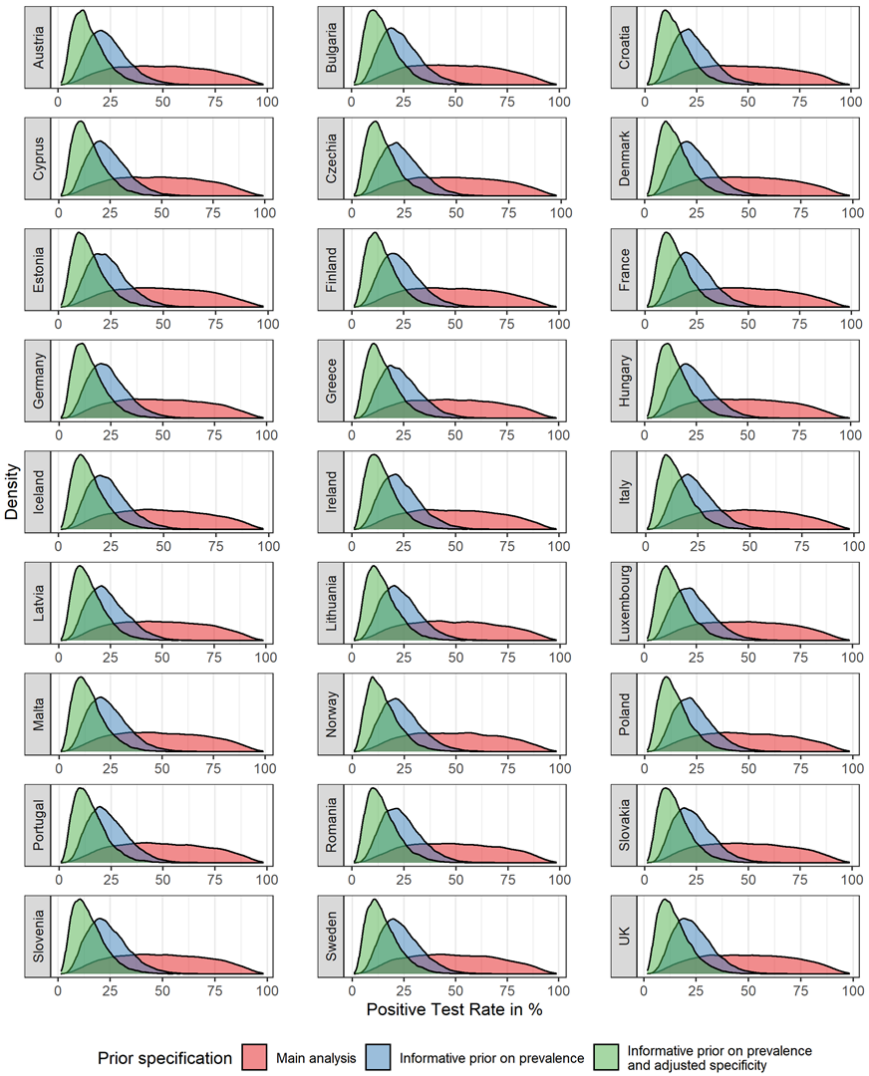
Some results of our re-analysis of the EHIS data (BMJ Mental Health)

Stan model for analyzing multiple countries

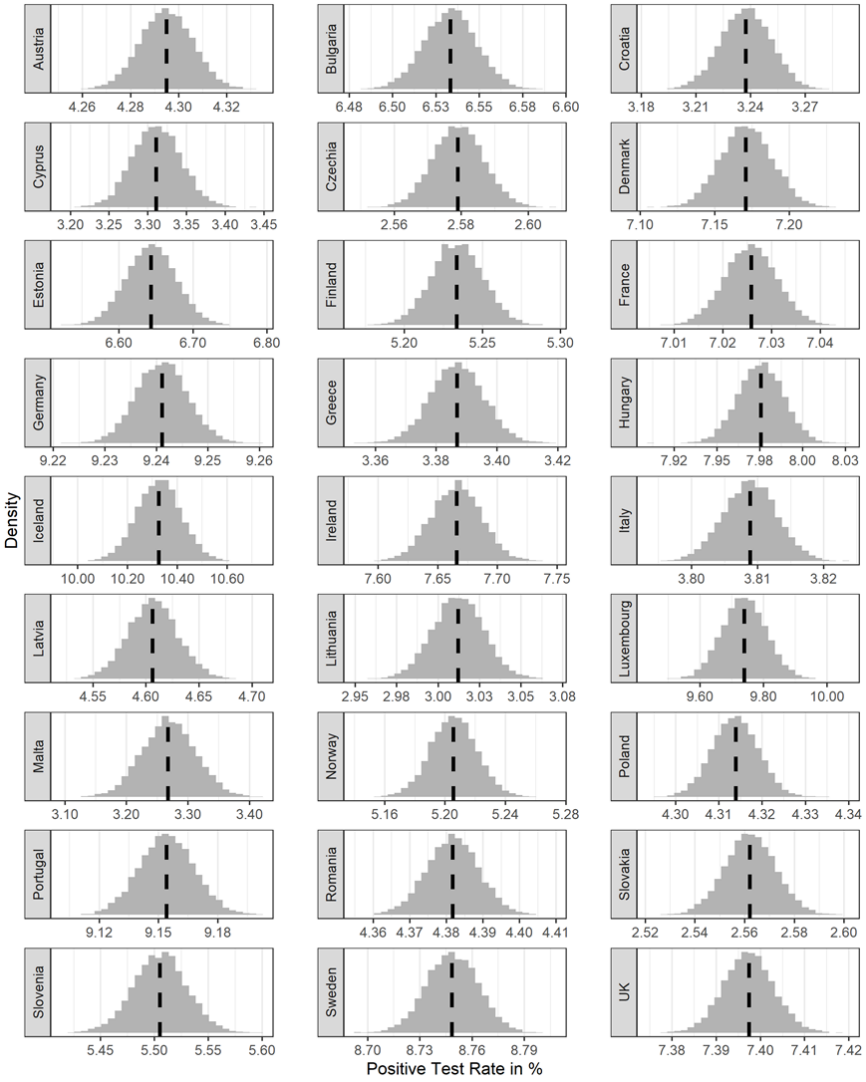
```
data {
  int<lower = 1> J; // number of countries
  int<lower = 0> n[J]; // number of tested persons
  int<lower = 0> y[J]; // number of positive persons
  vector[2] mu;
  vector<lower = 0>[2] prev_beta;
  cov_matrix[2] Sigma;
}
parameters {
  vector<lower = 0, upper = 1>[J] prev;
  matrix[J,2] logit_Se_Sp;
}
transformed parameters {
  vector[J] logit_Se = logit_Se_Sp[,1];
  vector[J] logit_Sp = logit_Se_Sp[,2];

  vector[J] Se = inv_logit(logit_Se);
  vector[J] Sp = inv_logit(logit_Sp);

  vector<lower = 0, upper = 1>[J] p;
  for(i in 1:J){
    p[i] = prev[i] * Se[i] + (1 - prev[i]) * (1 - Sp[i]);
  }
}
model {
  y ~ binomial(n, p);
  for(i in 1:J){
    logit_Se_Sp[i,] ~ multi_normal(mu, Sigma);
  }
  prev ~ beta(prev_beta[1],prev_beta[2]);
}

generated quantities {
  int y_pred[J] = binomial_rng(n, p);
}
```


Positive test rate
Iceland 10.3
Luxembourg 9.7
Germany 9.2
Portugal 9.2
Sweden 8.7
Hungary 8.0
Ireland 7.7
UK 7.4
Denmark 7.2
France 7.0
Estonia 6.6
Bulgaria 6.5
Slovenia 5.5
Finland 5.2
Norway 5.2
Latvia 4.6
Romania 4.4
Poland 4.3
Austria 4.3
Italy 3.8
Greece 3.4
Cyprus 3.3
Malta 3.3
Croatia 3.2
Lithuania 3.0
Czechia 2.6
Slovakia 2.6
Country
0.0% 2.5% 5.0% 7.5% 10.0% 12.5% 15.0%
Prevalence
Prior specification
Main analysis
Informative prior on prevalence
Informative prior on prevalence and adjusted specificity

Prior predictive checks



Posterior predictive checks



Conclusions

- Positive test rates are rarely true prevalences.
- Be critical when you hear about reported prevalences that appear surprisingly high (or low) and check whether they have considered diagnostic inaccuracy.

Specific for depression:

- Prevalence of depression is probably overestimated when positive rate of an imperfect diagnosis tool is used as prevalence estimator.
- The uncertainty is much larger than typically appreciated, even when using informative priors.
- The reported specificity estimates are likely to be too low
  - because when specificity is 86%, more people than observed would be diagnosed incorrectly with depression (because even with 0% prevalence, there should be 14% (false-)positives).

Learning objectives – achieved?

- Understand the structure of a simple BLCMs for prevalence estimation
- Familiarize oneself with the peculiarities of prevalence estimation in depression
- Comprehend how the prior was generated from meta-analysis

- Fit BLCM in Stan
- Adjust the priors for the BLCM and compare results

- Understand that the data contains information about specificity
- Understand that diagnostic studies do not provide perfect/appropriate prior information

References

- Fischer et al. (2023) Prevalence estimates of major depressive disorder in 27 European countries from the European Health Interview Survey: accounting for imperfect diagnostic accuracy of the PHQ-8. BMJ Ment Health 2023;26:e300675.

- Joseph et al. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. American Journal of Epidemiology, 141(3), 263- 272.

- Gelman & Carpenter (2020). Modeling a test with uncertain sensitivity and specificity. Journal of the Royal Statistical Society: Series C (Applied Statistics).

- Levis et al. (2020). Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis. Journal of Clinical Epidemiology, 122(6), 115-128.e1.

- Riley et al. (2008). Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Statistics in Medicine, 27, 6111–6136.