Bayesian statistics: basic concepts and Bayesian options in SAS

Mark Janssens (Grünenthal)

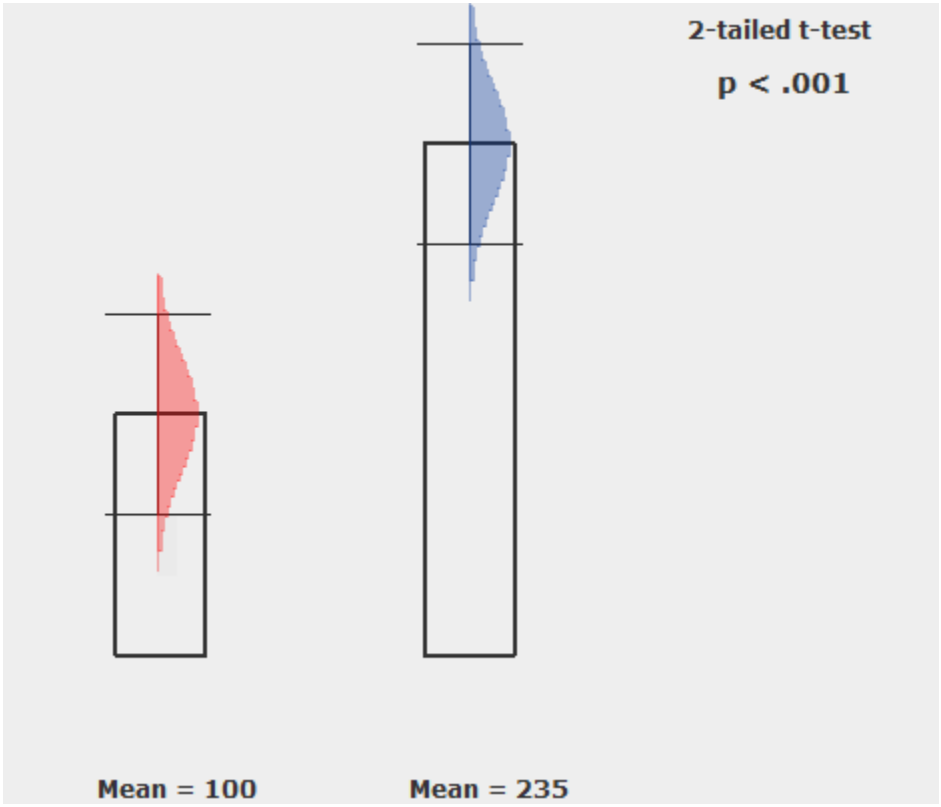
1. Basic concepts

Bayesian statistics

- Bayesian inference is a method of statistical inference in which evidence is used to update the uncertainties of competing probability models
(Wikipedia)

Frequentist statistical inference

- Treatment effect in ♀ ♂



- Treatment effect? Yes.
  Less than 0.1% to observe such a result if the null hypothesis of no difference is correct
  P (data | β) < 5% (two-sided)

Frequentist statistical inference

- Does not state anything about the SIZE of the treatment effect
- Distinguishes between statistical significance and clinical relevance
- Sees probability as a long-term frequency
  (if I repeat the trial 1000 times under the null-hypothesis, I only observe this result … times)
- Considers: P (data | β)
- Treatment effect

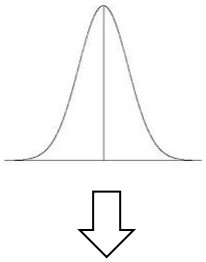
  $\tilde{\beta} \pm SE$

  confidence intervals

Bayesian statistical inference

- Does state something about the SIZE of the treatment effect
- Sees probability as a degree of (un)certainty about a parameter value
  (I am 95% percent certain that the treatment effect is greater than 0)
- Considers: P (β | data) or… P (β | data, β_0) with β_0 containing prior info on β
- Treatment effect

  $\tilde{\beta}$ + an entire distribution is estimated

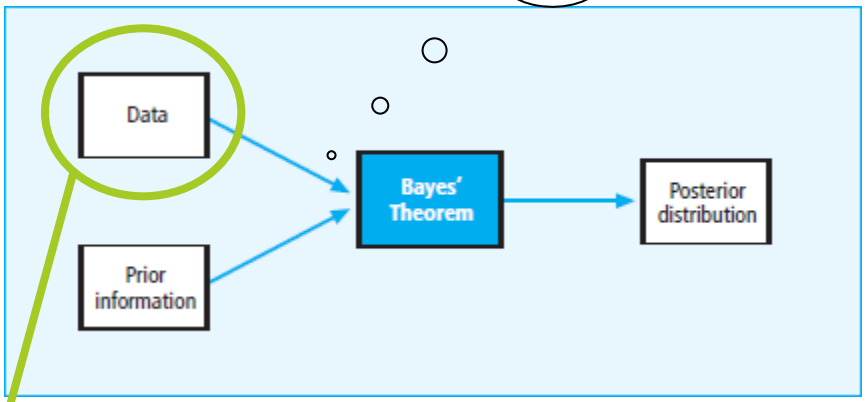


credibility intervals or posterior probability intervals

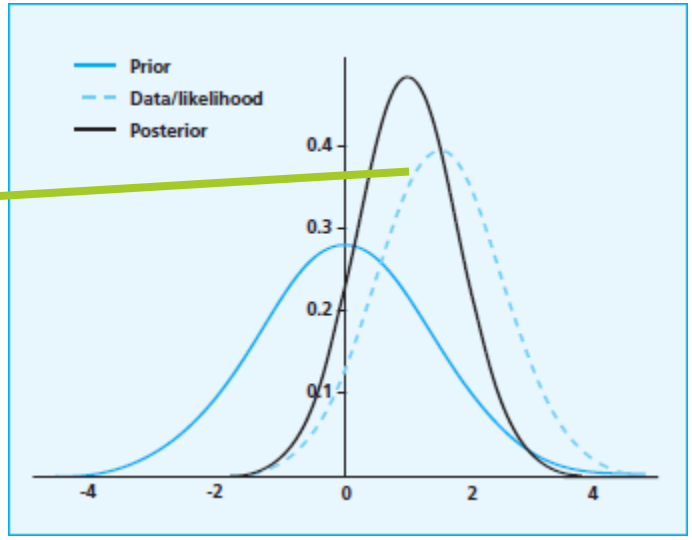
Bayesian statistics

- Can incorporate prior knowledge about β

competing probability models



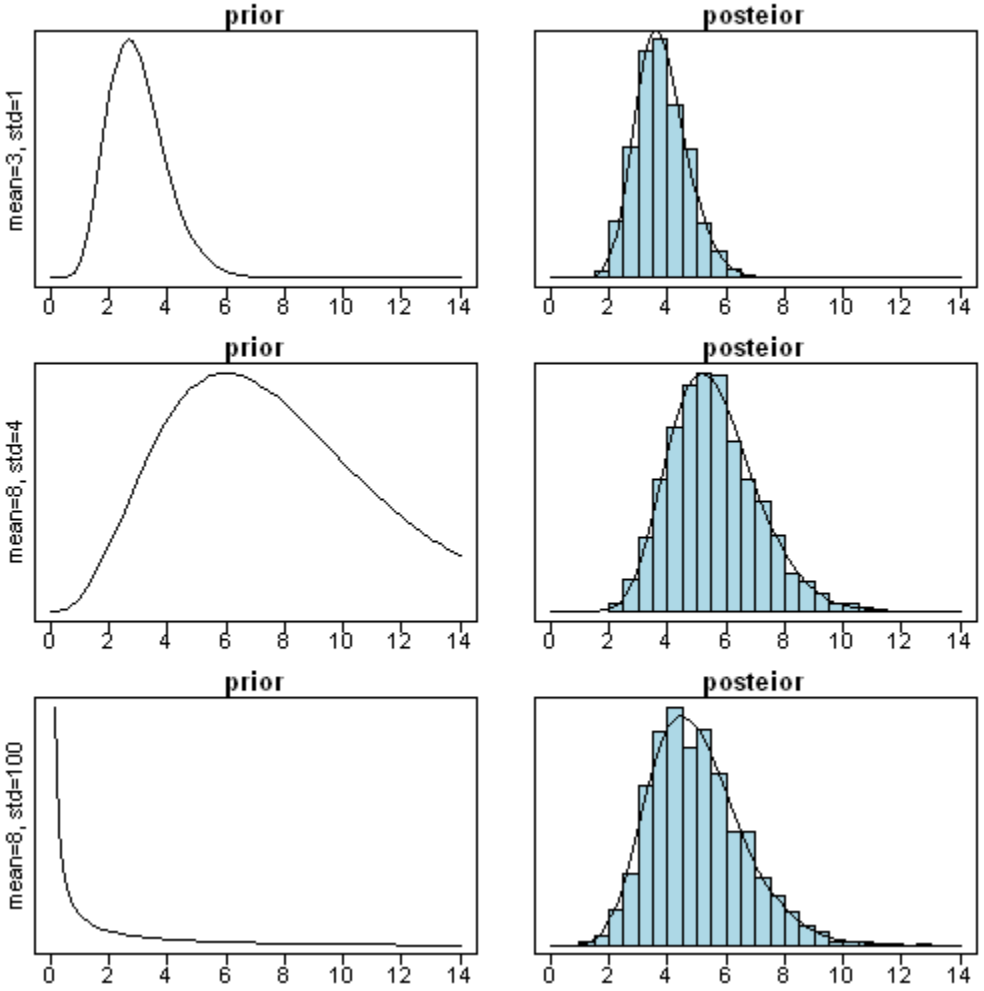
Frequentist statistics



Bayesian statistics

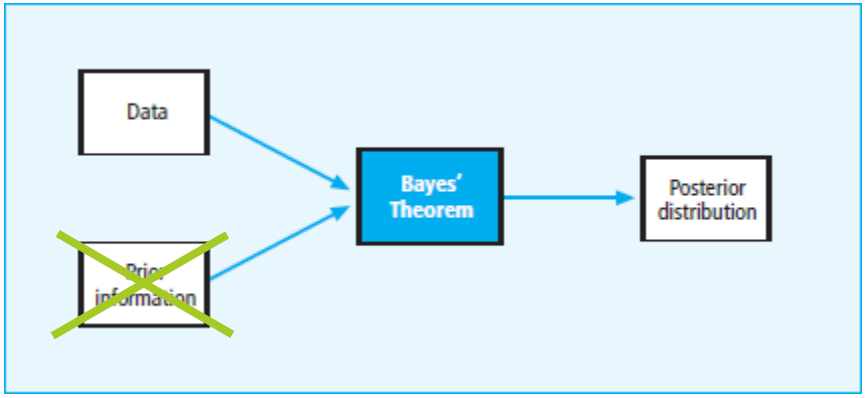
- Different priors lead to a different end result

competing probability models



Bayesian statistics – non-informative priors

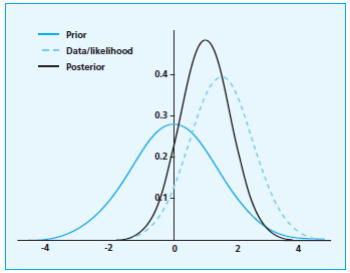
- Imagine the prior distribution is a horizontal line, this means:
  no specific range for β is suggested by the prior
- What would then be my best guess about β? What I observe in my data!
- $P (\beta \mid data, \beta_0) = P (\beta \mid data)$
  β_0 is not informative when it comes to estimating β
- There is no competition with the prior → the posterior will just be the likelihood
- Non-informative or flat prior



Bayesian statistics – conjugate priors

- Posterior distribution is a mix of prior distribution and data
- A posterior distribution is not necessarily a normal, binomial, … distribution
- However if the data is normal and the prior for μ is chosen normal → posterior for μ is a known distribution:

  normal prior for μ + normal data → normal posterior for μ

  Other examples:

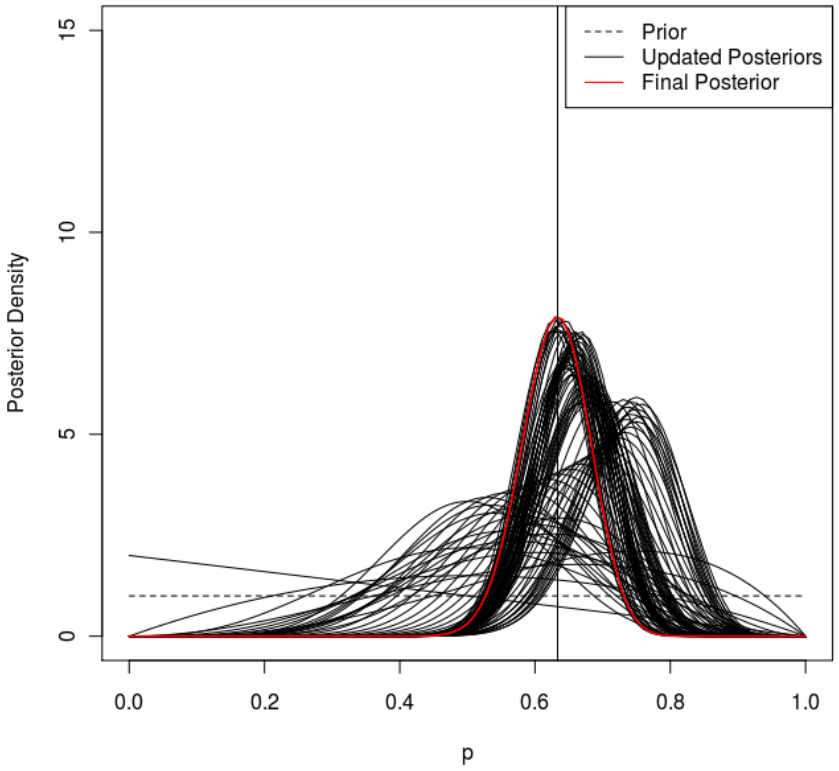
  inverse gamma prior for σ^2 + normal data → t-posterior for σ^2
  gamma prior for λ + Poisson data → gamma posterior for λ

- If prior and posterior are from the same family of distributions:
  - prior and posterior are then called conjugate distributions, and
  - prior is called a conjugate prior for the likelihood.

Bayesian statistics – conjugate priors

- The usefulness of a conjugate prior is that the corresponding posterior distribution will be in the same family, and the calculation may be expressed in closed form.

- In practice, iterative sampling (not closed form) is mostly used.



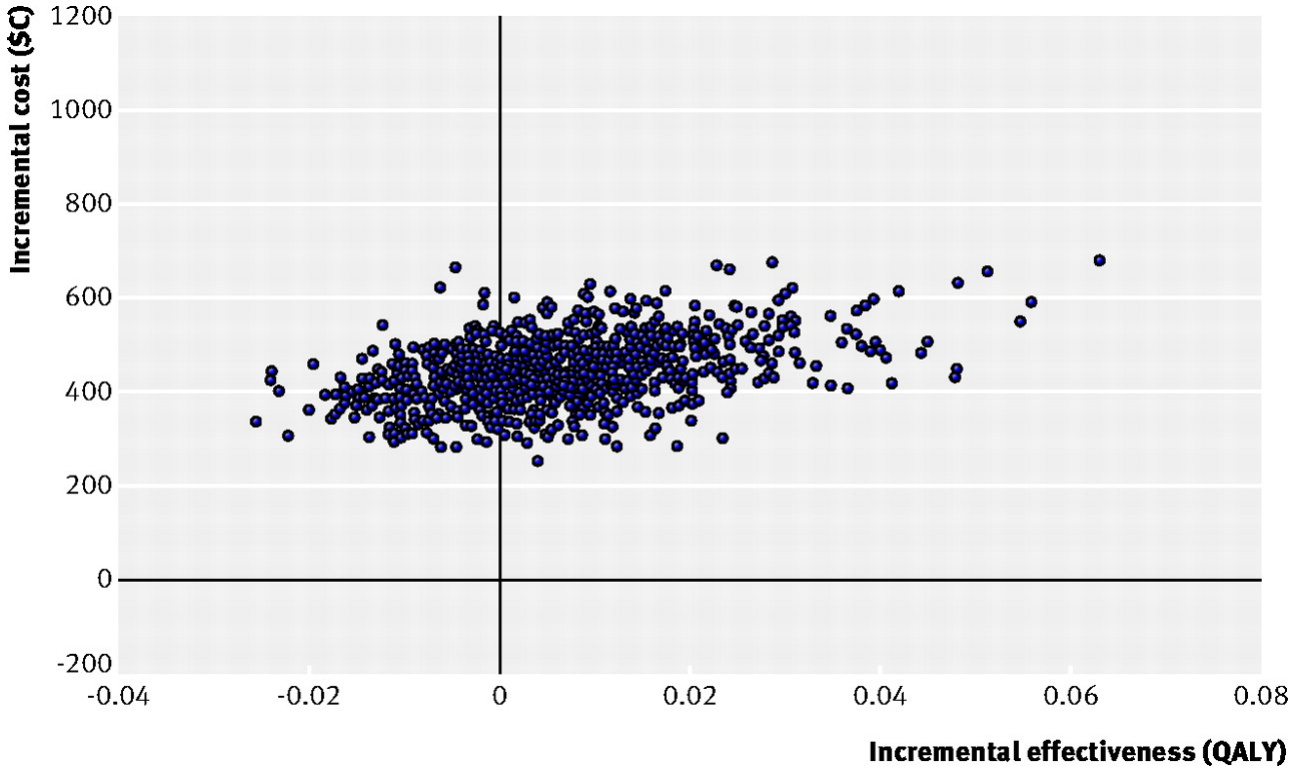
Bayesian statistics – examples

- Proof of concept trial
  - New drug, placebo, well-studied reference drug
  - Limit the number of subjects of reference drug (just for assay sensitivity)
  - Include at time of analysis trial data + prior knowledge regarding reference drug

- Health economics – probabilistic sensitivity analysis
  - Relation between drug effectiveness and drug cost is not one point estimate
  - Strategies for sensitivity analysis
    - Deterministic: change input parameters one at a time
    - Probabilistic:
      → run with all key parameters through plausible ranges
      → replace estimates of probabilities, utilities, and costs with probability distributions

Bayesian statistics – examples

- Health economics – probabilistic sensitivity analysis



Population based screening for chronic kidney disease: cost effectiveness study (BMJ 2010;341:c5869)

2. Bayesian options in SAS

Bayesian options in SAS

- Bayesian options in common PROCEDURES

	for
– PROC MIXED	Gaussian multilevel data
– PROC GENMOD	Gaussian, binary or Poisson data
– PROC PHREG	survival data
– PROC LIFEREG	survival data

- Bayesian procedure PROC MCMC

Example – data/likelihood

- Clinical trial
- 150 subjects
- 2 treatments → 75 subjects per treatment (1:1 ratio)
- Treatment effect is continuous measure
- 15 timepoints → 15 repeated measures

- What is the treatment effect at endpoint?

- Basic model for the sake of example:

$$Y_i = \beta_0 + b_{0,i} + \beta_1 \cdot Ybase + \beta_{2,k} \cdot Treatment + \beta_{3,jk} \cdot Treatment \cdot Time + \varepsilon_{ijk}$$

for subject i, timepoint j, treatment k

→ random intercept model with time as a factor

Example – data/likelihood

```
proc mixed data=mixed;
 class subjid arm week;
 model change = base arm arm*week;
 random intercept / subject=subjid;
 lsmeans arm*week / pdiff cl;
 ods output diffs=diffs1;
run;
```

→ Fixed effects

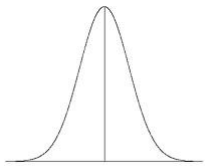
→ Random effect: random intercept

→ What is the treatment effect at endpoint?

- Treatment effect

Estimate	Standard Error
1.3637	0.4249

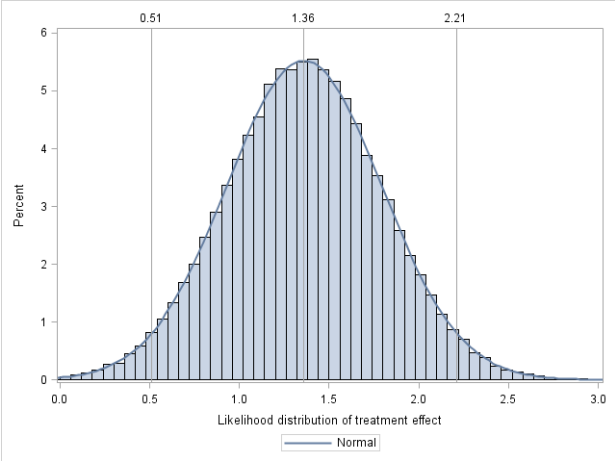
$\tilde{\beta} \pm SE$



let's go Bayesian

Example – flat prior

```
proc mixed data=mixed;
 class subjid arm week;
 model change = base arm arm*week;
 random intercept / subject=subjid;
 lsmeans arm*week / pdiff cl;
 ods output diffs=diffs1;
 prior flat / out=flat1 nsample=100000 seed=09052012;
 ods output base=basedensities1;
run;
```



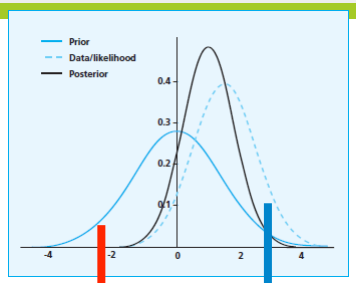
posterior mean = 1.36

should be the same as likelihood (flat prior)

The ‘win’ is the distribution to work with

Example I – informative prior regarding treatment effect

- Informative prior on mean with known variance:




```
proc mcmc data=flat1 thin=5 nbi=5000 nmc=25000 monitor=(beta0 beta1
sigma1 posterior_mean);
 parms beta0 beta1 sigma1;
 prior beta0 ~ normal(mean=1.5, var=0.1); * informative prior mean;
 prior beta1 ~ normal(mean=0, var=0.01); * data dependent mean;
 prior sigma1 ~ igamma(shape=0.01, scale=0.01); * data dependent var;
 prec0=1/0.1;
 prec1=1/sigma1;
 posterior_mean = ( prec0*beta0 + prec1*beta1 ) / (prec0 + prec1);
 model trt effect ~ normal(mean=beta1, var=sigma1);
run;
```

μ_0=1.5

μ = 1.4

- Treatment effect

```
Estimate
  1.4484
```

not 1.3637 (original PROC MIXED)

Example II – informative prior regarding intersubject variability

- Informative prior on intersubject variability:

- Difference with previous step:
  - Prior on distribution of random effect — *not on linear combination of fixed effects*
  - Prior on variance of this distribution — *mean of random effects distribution is 0*

- Where is this distribution of random effects?

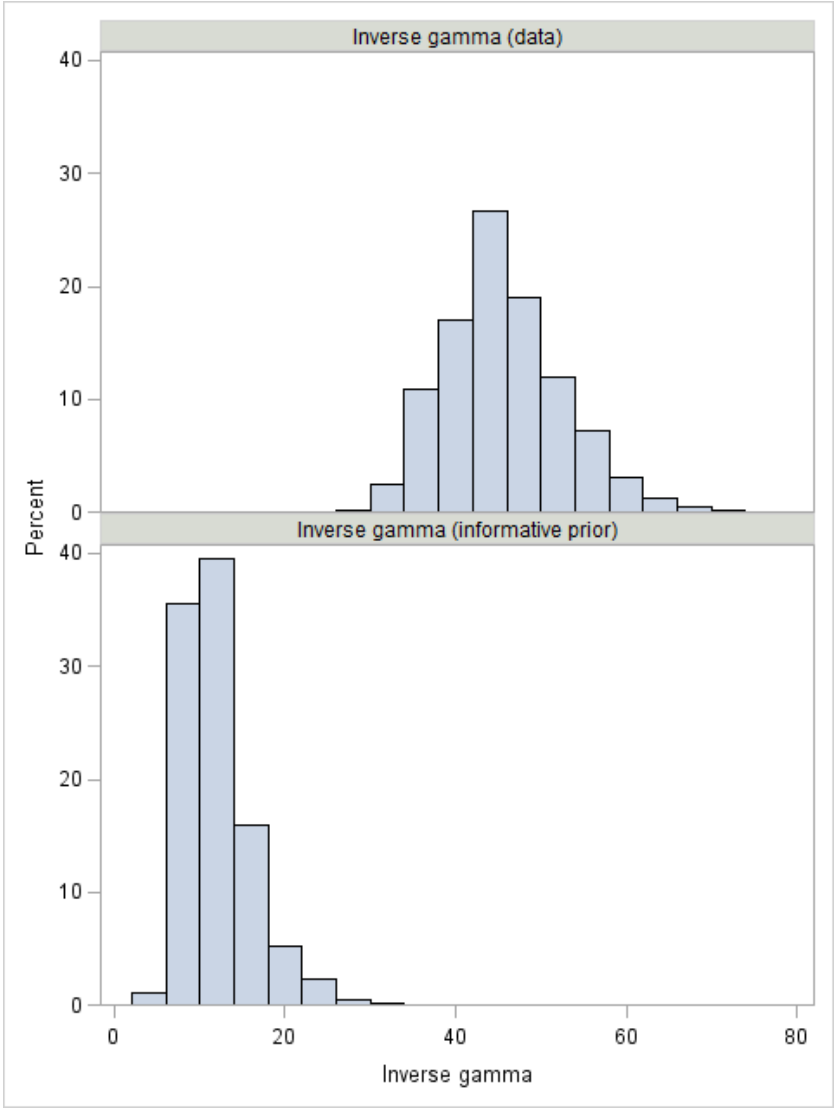
```
proc mixed data=mixed;
 class subjid arm week;
 model change = base arm arm*week;
 random intercept / subject=subjid;
 lsmeans arm*week / pdiff cl;
 prior flat / out=flat1 nsample=100000 seed=09052012;
 ods output base=basedensities1;
run;
```

→ Inverse gamma

Example II – informative prior regarding intersubject variability

- Let's narrow the intersubject variability

```
data basedensities2;
 set basedensities1;
 if density=1 then do;
  parm1=11;
  parm2=120;
 end;
run;
```



Example II – informative prior regarding intersubject variability

- Use the narrow distribution in PROC MIXED

```
proc mixed data=mixed;
 class subjid arm week;
 model change = base arm arm*week;
 random intercept / subject=subjid;
 lsmeans arm*week / pdiff cl;
 prior data=basedensities2 / out=info1 nsample=100000 seed=09052012;
run;
```

- Treatment effect

Estimate	Standard Error
1.3637	0.4045

not 0.4249 (original PROC MIXED)

Example I + II

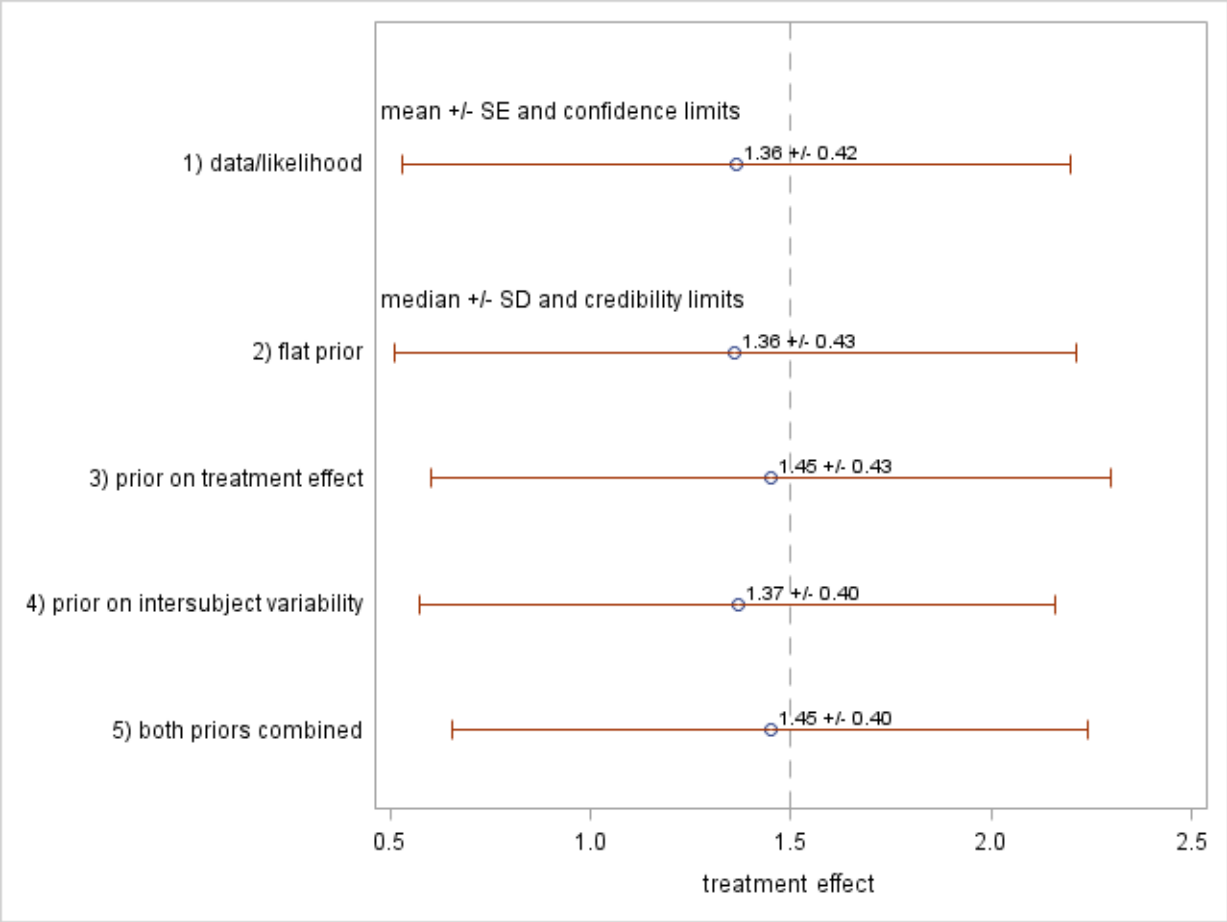
- Example I and II can easily be combined:

  informative prior regarding intersubject variability → feed this into PROC MIXED (Example I)

  informative prior regarding treatment effect → feed this into PROC MCMC (Example II)

- Alternative: fit the entire model in PROC MCMC
  (with SAS 9.3, using the RANDOM statement of PROC MCMC)

Example – overview



Concluding remarks

- Bayesian principles are *relatively* simple
- The number of Bayesian applications has grown rapidly
- Software companies follow this trend
- Implementation (coding + choice of priors) requires some experience

Thank you