A Case Study: Two-sample categorical data

Patrick Breheny

January 31

Introduction

- Today, we will look at a case study of real data an how a real analysis might proceed and be reported (it will also give us a chance to comment on some issues in categorical data analysis)
- Our example will be based on the Grampian Region Early Anistreplase Trial (GREAT), a randomized controlled trial in which 311 patients recovering from myocardial infarction were randomized to receive either anistreplase (a thrombolytic drug) or placebo
- In the study, 13 out of the 163 patients on anistreplase died, compared with 23 out of 148 in the placebo group

Likelihood

- The likelihood here is straightforward:

$$\text{Treatment: } Y_1 \sim \text{Binom}(163, \theta_1)$$
$$\text{Placebo: } Y_2 \sim \text{Binom}(148, \theta_2)$$

- The main questions for the analysis are:
  - Deciding on a quantity of interest
  - Choosing a prior for $\boldsymbol{\theta}$

Probability of a hypothesis using continuous priors

- To some extent, these two questions go hand in hand
- For example, suppose I decide that my quantity of interest is $p(\theta_1 = \theta_2|\mathbf{y})$; *i.e.*, the posterior probability that the risk of death with anistreplase is the same as it is with placebo
- If I choose continuous priors for θ_1 and θ_2, then I will obtain a continuous posterior, and can immediately conclude that $p(\theta_1 = \theta_2|\mathbf{y}) = 0$, regardless of what $\mathbf{y}$ actually is or what continuous priors I choose
- Clearly, this will not be helpful with regard to my quantity of interest

The mixture prior approach

- This is a good opportunity to discuss the idea of hypothesis testing from a Bayesian perspective
- Generally speaking, when it comes to testing the hypothesis that $\theta = \theta_0$ for a parameter of interest θ, there are two perspectives that one can take with regard to prior specification in Bayesian statistics
- If we believe (prior to seeing the data) that there is a meaningful probability that θ takes on an exact value θ_0, then we must reflect this in the prior
- Typically, this will involve some sort of a mixture prior, as the alternative hypothesis rarely specifies exact values for θ

A mixture prior for anistreplase

- In the anistreplase example, we might be interested in the quantity $\omega = \theta_2 - \theta_1$, a measure of the benefit provided by the drug, and therefore in testing the hypothesis that $\omega = 0$
- If we believe that there is, say, a 50% chance that ω is exactly 0, then our prior needs to reflect that
- On the other hand, if anistreplase is not identical to placebo, there is a broad range of benefits or harms that it might be associated with
- We have seen already seen an example of this with regard to binomial data for a crossover study in the 1-17 notes

Implications of a continuous prior

- The other perspective is the belief that θ follows some strictly continuous distribution of possible values
- If so, then any question of whether it is exactly equal to any one particular value θ_0 is meaningless; the only meaningful questions involve intervals of credible values for θ
- Applying this perspective to our example, the benefit of anistreplase as measured by $\omega = \theta_2 - \theta_1$ follow a continuous distribution and it is pointless to ask about the probability that $\omega = 0$

A continuous prior for anistreplase

- It would be meaningful, however, to consider the probability that anistreplase provides a benefit: $p(\omega > 0)$
- Or, perhaps we would consider it unimportant if anistreplase affects the probability of death by less than 3%
- We might be interested in considering $p(\theta \in [-0.03, 0.03])$; the region $[-0.03, 0.03]$ is sometimes referred to as the "region of practical equivalence"

My choices for the GREAT study

- Of course, this would not be an appropriate measure for this study – if anistreplase dropped the risk of death from 4% to 1%, we would certainly say that it has a large, clinically significant effect
- For this reason, one typically looks at ratios such as the *relative risk*, θ_2/θ_1, or the *odds ratio*:

$$\omega = \frac{\theta_2/(1-\theta_2)}{\theta_1/(1-\theta_1)}$$

  as the quantity of interest
- I will proceed with this analysis choosing continuous priors (I feel that there is zero probability that anistreplase has *exactly* the same risk as placebo) and the odds ratio as my quantity of interest, although there are certainly other defensible choices

Community of priors

- We now turn to the choice of prior
- In particular, we will consider 4 possible priors:
  - Reference: $\theta_j \sim \text{Unif}(0, 1)$
  - Skeptical: A prior that is doubtful about the benefits of treatment
  - Optimistic: Opposite of the skeptical prior
  - Clinical: An informative prior based on past data

Reparameterization

- For the informative priors, it will be useful to specify priors directly in terms of the odds ratio, or more precisely, in terms of $\delta = \log \omega$ (logarithms are typically preferred when working with ratios, as it tends to make them roughly normally distributed)
- Letting $\alpha = \frac{1}{2}(\text{logit}\,\theta_1 + \text{logit}\,\theta_2)$ denote the average log-odds, where $\text{logit}(\theta) = \log\{\theta/(1-\theta)\}$, we have

$$\text{logit}\,\theta_2 = \alpha + \frac{1}{2}\delta$$

$$\text{logit}\,\theta_1 = \alpha - \frac{1}{2}\delta$$

- α is a nuisance parameter and will be given a relatively vague distribution, $\sim N(0, 1.6^2)$, by both the skeptical and optimistic prior

Skeptical prior

- A reasonable expression of doubt for the skeptical prior might be that anything beyond a two-fold change in the odds ratio is quite unlikely
- Specifically, if the skeptic believes that there is only a 5% chance that the true effectiveness of the treatment lies outside this range, we would have

$$\delta \sim N(0, \sigma_\delta^2)$$
$$\sigma_\delta = \frac{\log 2}{1.96}$$

Optimistic prior

- An optimistic prior, on the other hand, might believe that the most likely scenario is that anistreplase cuts the odds in half, and that there is only a small chance (5%) that it is harmful:

$$\delta \sim N(\log 2, \sigma_\delta^2)$$
$$\sigma_\delta = \frac{\log 2}{1.645}$$

- The art of asking questions to quantify prior beliefs and map them to specific priors is known as prior *elicitation*
- There is a large literature on this subject, and we are not truly doing it justice here, but focusing on believable ranges (as opposed to trying to directly specify distributional parameters) has been shown to be an effective strategy

Clinical prior

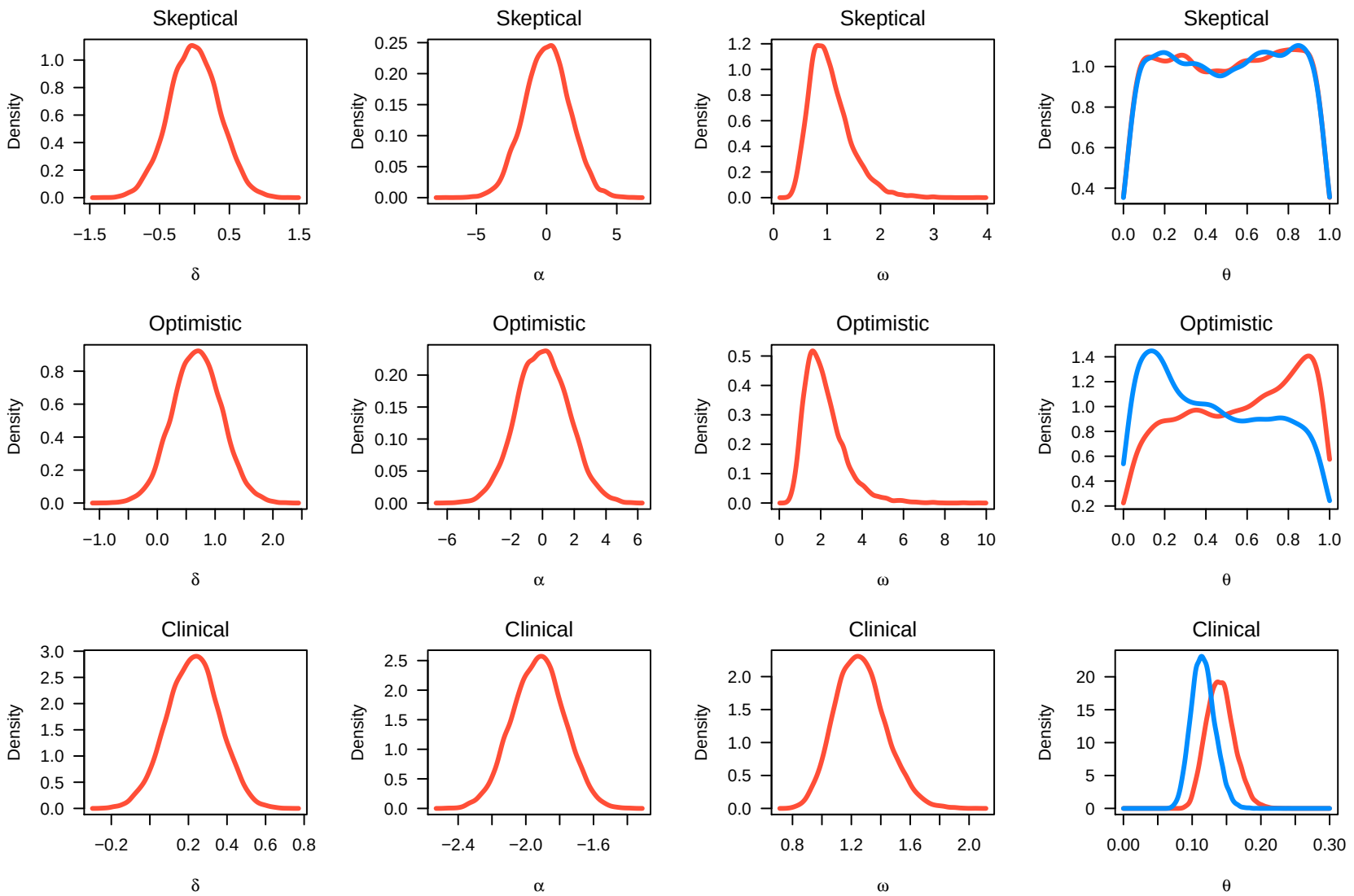
- Now suppose a clinician's informed opinion, based on a general familiarity with this condition, is that the odds of death in the placebo group is probably around 14%, and rather unlikely to be lower than 11% or higher than 18%
- Meanwhile, based on the results of a study of anistreplase conducted by a different group in a slightly different setting, it is believed that anistreplase likely lowers the odds of death by about 20%, and it is rather unlikely that its benefit is either 0 or 40%

Clinical prior (cont'd)

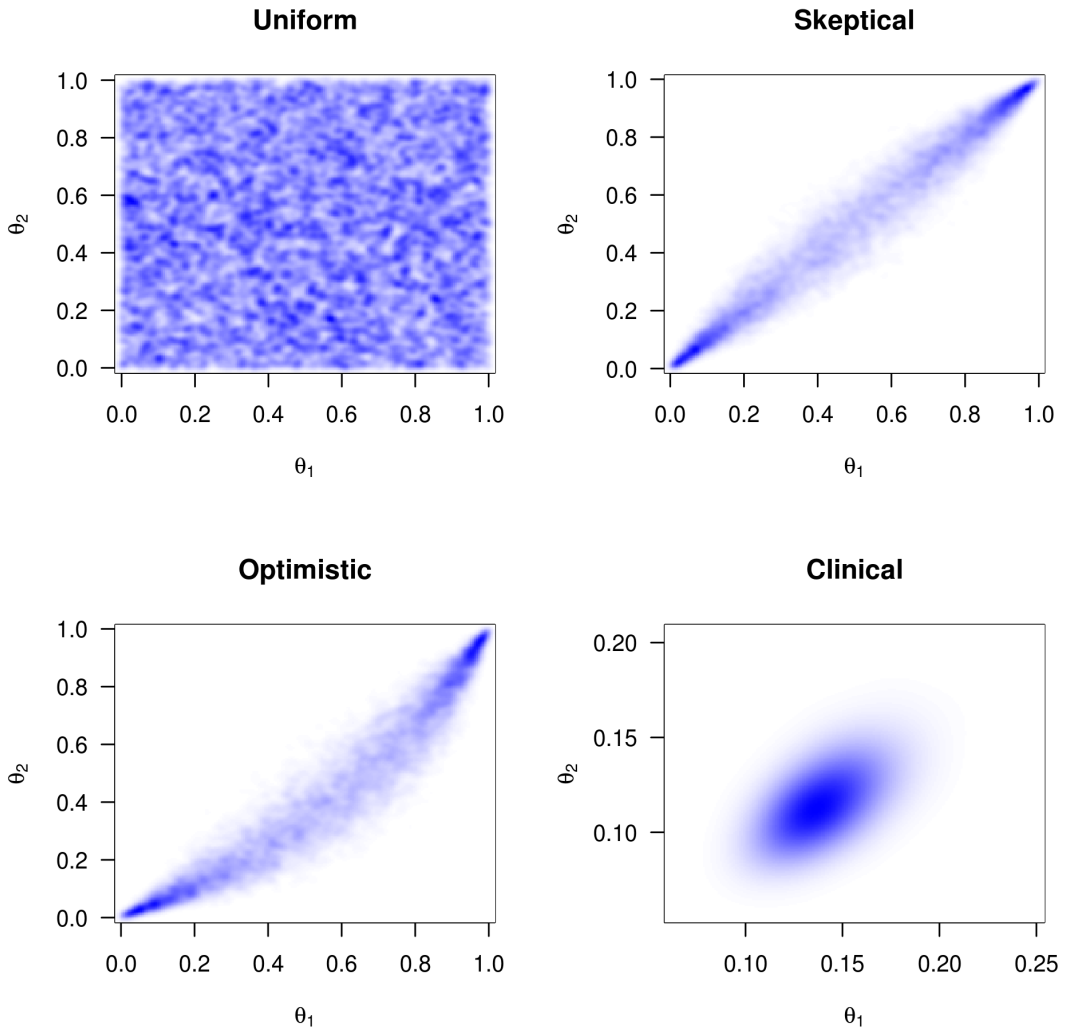
Taking "rather unlikely" to mean 10% probability, this would translate into priors:

$$\mu_\delta = \log(1/.8)$$
$$\sigma_\delta = \frac{\log(1/.8)}{1.645}$$
$$\mu_\alpha = \text{logit}(.14) - 0.5\log(1/.8)$$
$$\sigma_\alpha = \sqrt{\left(\frac{\text{logit}(.14) - \text{logit}(.11)}{1.645}\right)^2 - \frac{\sigma_\delta^2}{4}}$$

The priors


Skeptical
Density
δ
Skeptical
Density
α
Skeptical
Density
ω
Skeptical
Density
θ
Optimistic
Density
δ
Optimistic
Density
α
Optimistic
Density
ω
Optimistic
Density
θ
Clinical
Density
δ
Clinical
Density
α
Clinical
Density
ω
Clinical
Density
θ

Priors for θ_1 and θ_2 are dependent



Reporting of Bayesian analyses

- Now is as good a time as any to comment on how to report a Bayesian analysis
- In many ways, reporting Bayesian analyses and frequentist analyses are similar, but there are also a number of important differences to be aware of, as well as issues that come up in Bayesian analyses (such as priors) that do not come up in Frequentist analyses
- The guidelines we discuss here are appropriate for general reporting of Bayesian analyses for a journal article, but also good guidelines for reporting Bayesian analyses for this class in future homework assignments

Articles

We will include some examples drawn from the following articles:

- Bergamaschi et al. (2001). Predicting secondary progression in relapsingremitting multiple sclerosis: a Bayesian analysis. *Journal of the Neurological Sciences*, 189: 1321
- Hamilton (2001). Estimating treatment effects in randomized clinical trials with non-compliance: The impact of maternal smoking on birthweight. *Health Economics*, 10: 399-410
- Cancré et al. (2000). Bayesian analysis of an epidemiologic model of Plasmodium falciparum malaria infection in Ndiop, Senegal. *American Journal of Epidemiology*, 152: 760-770.

Essential components

First, the "essential" items – any Bayesian analysis must report at least these details:

- Description of model: Likelihood
- Description of model: Priors
- MCMC/Analytic details (how you calculated/sampled from the posterior)
- Report and interpret the posterior distribution

Description of model: Likelihood

- The description of the model for observable parameters would typically be in a "Methods" section; the length is dictated by the complexity of the model
- A simple model could be described in words, possibly with a short accompanying equation:

receiving encouragement and choosing treatment are either compliers or always-takers, while those who choose the treatment without encouragement are always-takers or defiers.

Let $p_i^k = \Pr(C_i = k)$ be the probability that subject i is of compliance type k, $k = c, d, n, a$, and $y_i^k(S_i, r_i)$ denote the outcome for subject i if she is of type k. Outcomes are normally distributed with

$$y_i^k(S_i, r_i) \sim N(\mu_r^k, \sigma_r^k) \tag{4}$$

The classifications in Table 1 imply that the econometric framework requires estimation of a finite mixture model over compliance type. From this framework, an estimate of the complier average treatment effect described in Equation (2) is given by the quantity $T^c = \mu_1^c - \mu_0^c$.

Description of model: Likelihood (cont'd)

A figure could also be used to illustrate the model:

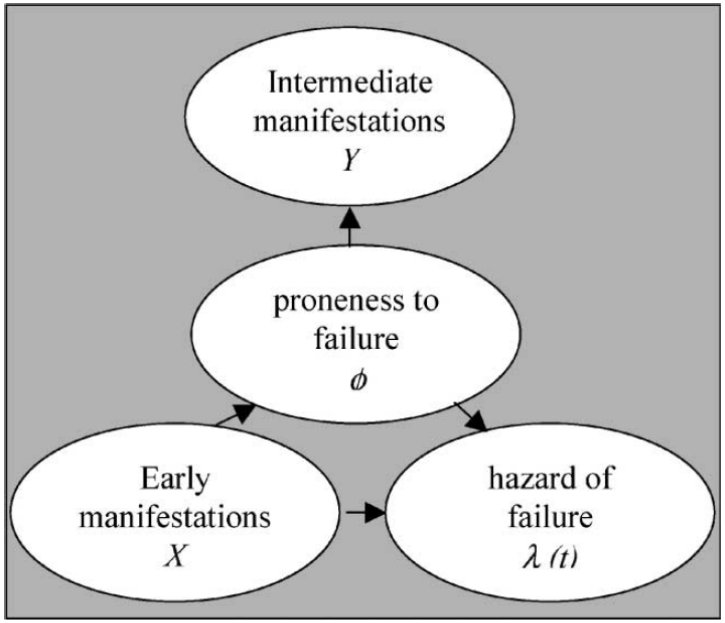


Fig. 1. Main assumptions underlying our model of disease course, represented in the form of a directed graph.

Appendix A. Statistical appendix

While the graph in Fig. 1 describes the conditional independence assumptions underlying our analysis, it does not provide a probabilistic interpretation of the data. To construct a full probability model of the problem, it is necessary to provide a quantitative specification of the way the nodes $\lambda(t)$, ϕ and Y depend on their respective "parent nodes" in the graph. For example, dependence of the node $\lambda(t)$ on their parent nodes, X and ϕ, and dependence of ϕ on X, were modeled by:

$$\lambda(t) = \lambda_0(t)\exp\left(\phi + \sum_k \delta_k X_k\right)$$

$$\phi \sim N\left(\sum_k \beta_k X_k, \sigma^2\right), \qquad (1)$$

where $\lambda_0(t)$ denotes baseline hazard function, the symbol $\sim$ stands for "distributed as" and $N(a, b)$ denotes Gauss-

In particular, if any important parameters of interest with special meaning are present in the model, the analysis should also discuss their interpretation

Description of model: Prior

- The model for the observable parameters is only part of a Bayesian model, of course: the prior must also be described
- Again, for simple uninformative priors, a short verbal description may be all that is needed:

Bayesian implementation

The first step in a Bayesian model is the formation of prior distributions for the parameters of the model, $\{\mu_r^k, \sigma_r^k, p^k\}$. Following standard practice, the prior on μ_r^k is assumed to be Gaussian with $N(\mu_0, M_0)$. The prior on the variance term σ_r^k is assumed to follow an inverse gamma distribution. Dirichlet process priors are assumed for the compliance probabilities.

Description of model: Prior (cont'd)

For informative priors, the prior needs not only to be described, but justified as well:

TABLE 2. Prior distributions adopted for the model parameters in a study of *Plasmodium falciparum* malaria, Ndiop, Senegal, 1993–1994

Parameter	Sources	Distribution	Mean	Standard deviation	Minimum	Maximum
r_1	Molineaux and Gramiccia (9) and Nedelman (21)	Lognormal	0.00118*	5†		
r_2	Molineaux and Gramiccia (9) and Nedelman (21)	Lognormal	0.0134*	5†		
α_1	Molineaux and Gramiccia (9) and Nedelman (21)	Lognormal	0.0108*	5†		
α_2	Molineaux and Gramiccia (9) and Nedelman (21)	Lognormal	0.00026*	5†		
τ	Molineaux and Gramiccia (9), Earle et al. (11), and Anderson and May (12)	Truncated lognormal	365*	2†	90	1,095
N_1	Gilles and Warrell (4) and Molineaux and Gramiccia (9)	Truncated lognormal	15*	2†	5	50
f	—‡	Uniform	0.5	0.289	0	1
b_1	—‡	Uniform	$(1 + b_2)/2$	$(1 - b_2)/\sqrt{12}$	b_2	1
b_2	—‡	Uniform	$f/2$	$f/\sqrt{12}$	0	f
$X_3(0)$	—‡	Uniform	0.5	0.289	0	1
$F_{Y_3(0)}$	—‡	Uniform	0.5	0.289	0	1
$F_{Y_2(0)}$	—‡	Uniform	0.5	0.289	0	1
$z_1(0)$	—‡	Uniform	50	28.9	0	100
$z_2(0)$	—‡	Uniform	50	28.9	0	100
$z_3(0)$	—‡	Uniform	50	28.9	0	100

* Geometric mean.
† Geometric standard deviation (exponential of the standard deviation in log space).
‡ Given the lack of prior information, an uninformative uniform prior was used.

MCMC/analytic details

- An analysis must also provide some details about how the posterior was sampled from (or how it was calculated directly, if that was possible)
- These include the software used, the number of samples, and various convergence diagnostics that we haven't discussed yet
- An example:

> cent (40). The convergence of several Markov chains to $p(\boldsymbol{\theta}|\mathbf{D})$ was assessed using Gelman and Rubin's $\hat{R}$ diagnostic (41). The parameter sets recorded after equilibrium was reached were used to form histograms or compute summary statistics of the posterior distributions for estimands of interest (e.g., marginal parameter distributions, combinations of parameters, or model predictions). Obtaining the posterior distribution of model predictions required running the malaria model once for each parameter set recorded. All of the above computations were performed using MCSim software, version 4.2 (34).

Posterior and its interpretation

- Finally, the posterior distribution for any quantities of interest must be reported and interpreted
- In particular, it is crucial to report on both the central tendency (posterior mean, median, or mode) of the distribution, as well as to report the spread of the distribution (posterior SD, quantiles, or using a credible interval)
- Of course, if the analysis centers on one quantity of particular interest, it may be worth plotting the entire posterior

Hypothesis probabilities

- Note that if we used a mixture prior with point masses on certain "null" values or hypotheses, then we would also report the posterior probabilities for those hypotheses (we will discuss a related quantity called the *Bayes factor* later in the course)
- It is worth noting, however, that most Bayesian analyses do not involve such results
- This is in stark contrast to frequentist statistics, where special consideration of probabilities pertaining to null values (*i.e.*, the p-value) is ubiquitous
- Whether the lack of p-value-like results in Bayesian statistics is a drawback or an advantage depends on one's perspective; regardless, it may take some getting used to when it comes to reporting results

Reporting the posterior: Example 1

Table 2
Estimates of local relative risks (LRRs) associated with early predictors (observed at 1 year after disease onset)
For each individual predictor, the table reports: posterior means of LRR (mean LRR); mean log LRR; 95% Bayesian credible interval for the LLR.

	Mean LRR	Mean log LRR	95% CI
Age at onset	1.05	0.05	1.02–1.09
Female gender (117, 63%)	0.39	−1.07	0.17–0.78
Sphincter onset (17, 9%)	2.98	0.93	1.10–6.10
Pure motor onset (22, 21%)	2.11	0.62	0.90–4.20
Motor-sensory onset (54, 29%)	2.40	0.81	1.15–4.41
Sequelae after onset (44, 23%)	1.76	0.52	1.04–2.88
No. of involved FS at onset	1.39	0.32	1.16–1.64
No. of sphincter plus motor relapses	2.10	0.71	1.56–2.89
EDSS ≥ 4 outside relapse (6, 3%)	2.28	0.44	0.40–6.50

Reporting the posterior: Example 2

TABLE 3. Summary of the posterior (fitted) distributions for the model parameters in a study of *Plasmodium falciparum* malaria, Ndiop, Senegal, 1993–1994

Parameter	Median	Mean (SD*)	Geometric mean (GSD*)	2.5th percentile	25th percentile	75th percentile	97.5th percentile
r_1	0.0049	0.030 (0.18)	0.0047 (3.3)	0.0005	0.0026	0.0082	0.057
r_2	0.053	0.055 (0.015)	0.054 (1.3)	0.033	0.045	0.063	0.095
α_1	0.12	0.33 (0.51)	0.12 (4.9)	0.0055	0.04	0.39	2.0
α_2	0.0047	0.0047 (4.6×10^{-4})	0.0047 (1.1)	0.0037	0.0044	0.0051	0.0055
τ	230	235 (30)	230 (1.1)	180	220	260	290
N_1	22	22.5 (2.9)	22 (1.2)	15	21	24	27.5
f	0.030	0.038 (0.051)	0.032 (1.5)	0.019	0.026	0.035	0.076
b_1	0.39	0.43 (0.16)	0.41 (1.4)	0.26	0.33	0.48	0.91
b_2	0.024	0.024 (0.005)	0.023 (1.2)	0.015	0.021	0.027	0.035
$X_3(0)$	0.66	0.62 (0.26)	0.525 (2.0)	0.080	0.42	0.84	0.98
$F_{Y_3(0)}$	0.37	0.42 (0.28)	0.30 (2.7)	0.025	0.18	0.66	0.96
$F_{Y_2(0)}$	0.42	0.45 (0.27)	0.33 (2.6)	0.030	0.22	0.66	0.95
$z_1(0)$	51	50.5 (28)	39 (2.5)	3.7	27	74	96
$z_2(0)$	48	49 (28)	37 (2.5)	3.3	25	72	96
$z_3(0)$	52	51 (28)	38 (2.6)	2.7	27	75	97

* SD, standard deviation; GSD, geometric standard deviation.

Reporting the posterior: Example 3

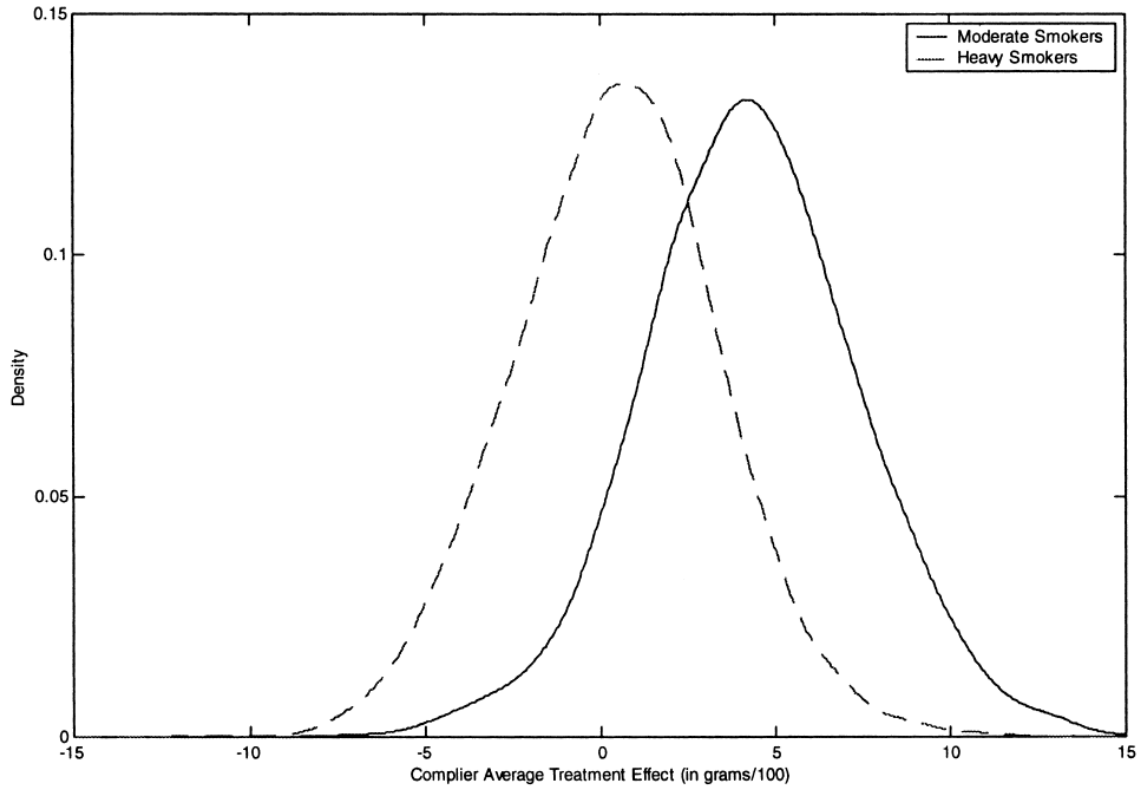


Figure 3. Posterior distributions of complier average treatment effect (T^c) for moderate and heavy smokers prior to pregnancy

Health Econ. **10**: 399–410 (2001)

Optional components

There are also a number of optional components that may be important to include for some analyses but not others, such as:

- Motivating the use of Bayesian methods
- Robustness to the specification of priors
- Model diagnostics (as opposed to MCMC diagnostics)
- Making the code or MCMC output available

Description of the model

For our analysis of the GREAT study, we could describe the model using these equations, with an accompanying verbal descriptions, a verbal description of the uninformative/reference prior, and all of the justifications of the priors listed earlier:

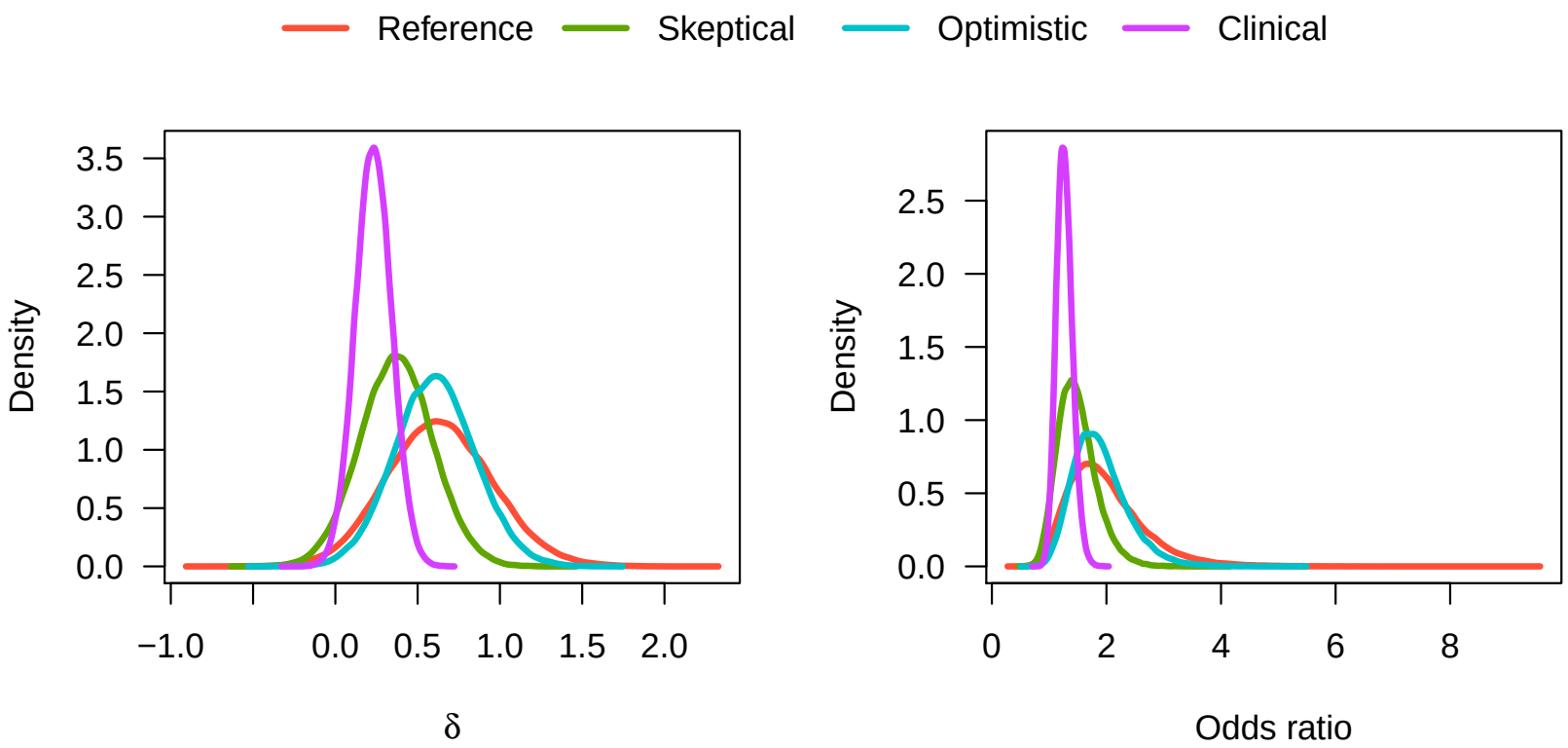
$$\begin{aligned} Y_i|\theta_i &\sim \text{Binom}(\theta_i, n_i) \\ \text{logit}\,\theta_i &= \alpha \pm \delta/2 \\ \alpha &\sim \text{N}(\mu_\alpha, \sigma_\alpha^2) \\ \delta &\sim \text{N}(\mu_\delta, \sigma_\delta^2) \end{aligned}$$

Analytic details

- With this analysis, the following description of the MCMC details would be sufficient: "For each prior, 100,000 values of $\{\theta_1, \theta_2\}$ were drawn from their posterior distribution. JAGS (Plummer 2003) was used to carry out the MCMC sampling."
- Given the relative simplicity of this analysis, I doubt anyone would question the validity of the MCMC sampling, but this is an important consideration with more complicated models; we will discuss diagnostics pertaining to these considerations in a few weeks

Reporting the posterior

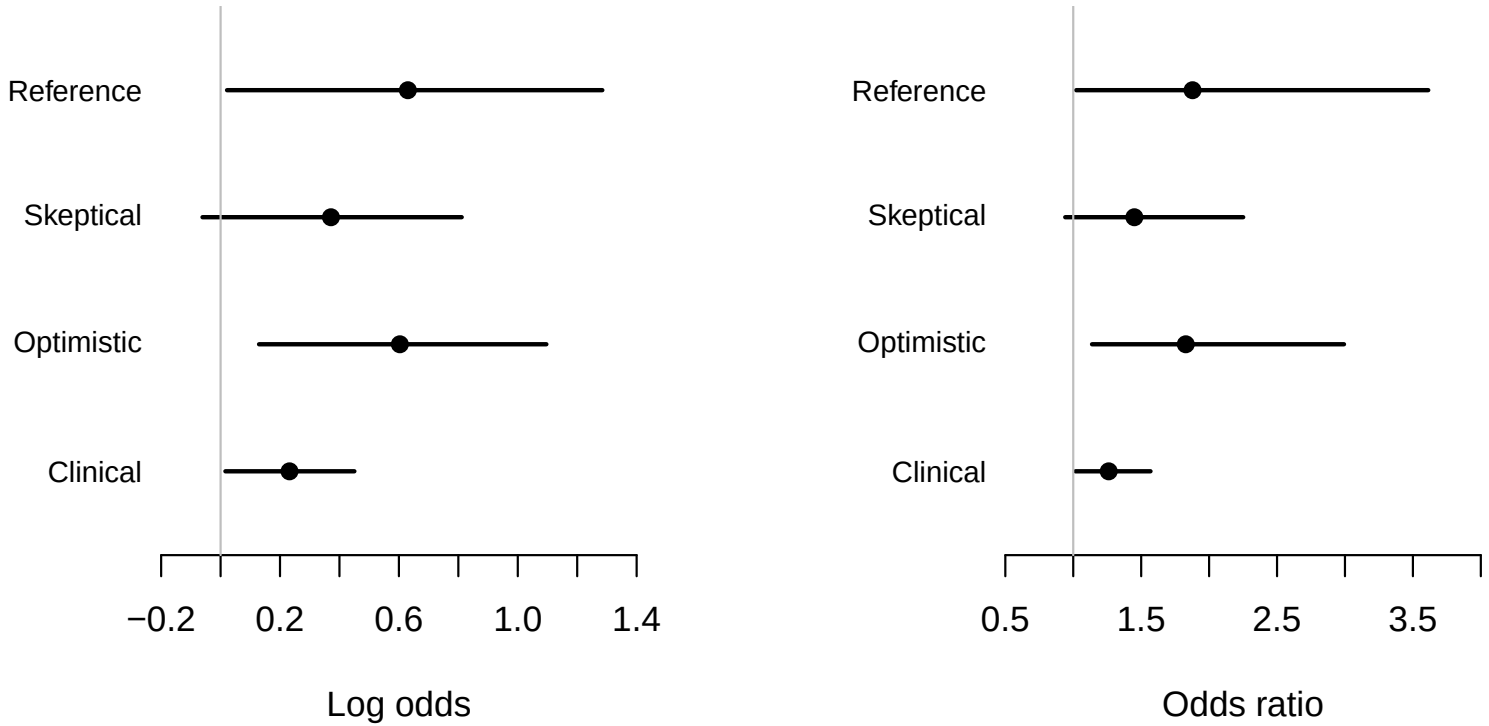
The posteriors:



In an actual journal article, I might not include this, only the summary on the next slide

Reporting the posterior (cont'd)

Summarizing the posteriors with posterior mode, 95% central interval:



Conclusions

- Note that our conclusions here depend quite a bit on our priors: an optimistic or completely uninformed individual would be fairly convinced by the study that anistreplase has a beneficial effect and reduces the risk of death following myocardial infarction
- For a skeptical individual, this trial would provide some grounds for optimism, but would not be sufficient to convince him/her of anistreplase's benefit – there is a considerable posterior probability that the benefit is near zero
- Finally, for a clinical opinion heavily informed by past studies and experience, this trial, being fairly consistent with prior expectations, would not substantially alter belief concerning anistreplase's benefit

Commentary

- Many people feel uncomfortable with this notion of conclusions being affected by prior beliefs
- However, the question of how to weigh new evidence in light of other research pertaining to anistreplase is obviously a very important question, and it is a question that frequentist statistics doesn't really help us with
- In a non-Bayesian analysis, we would be left to consider this question, typically in the discussion, using qualitative, opaque, vague, and possibly flimsy arguments
- The Bayesian approach makes the weighing of outside evidence quantitative, overt, and rigorous

Commentary (cont'd)

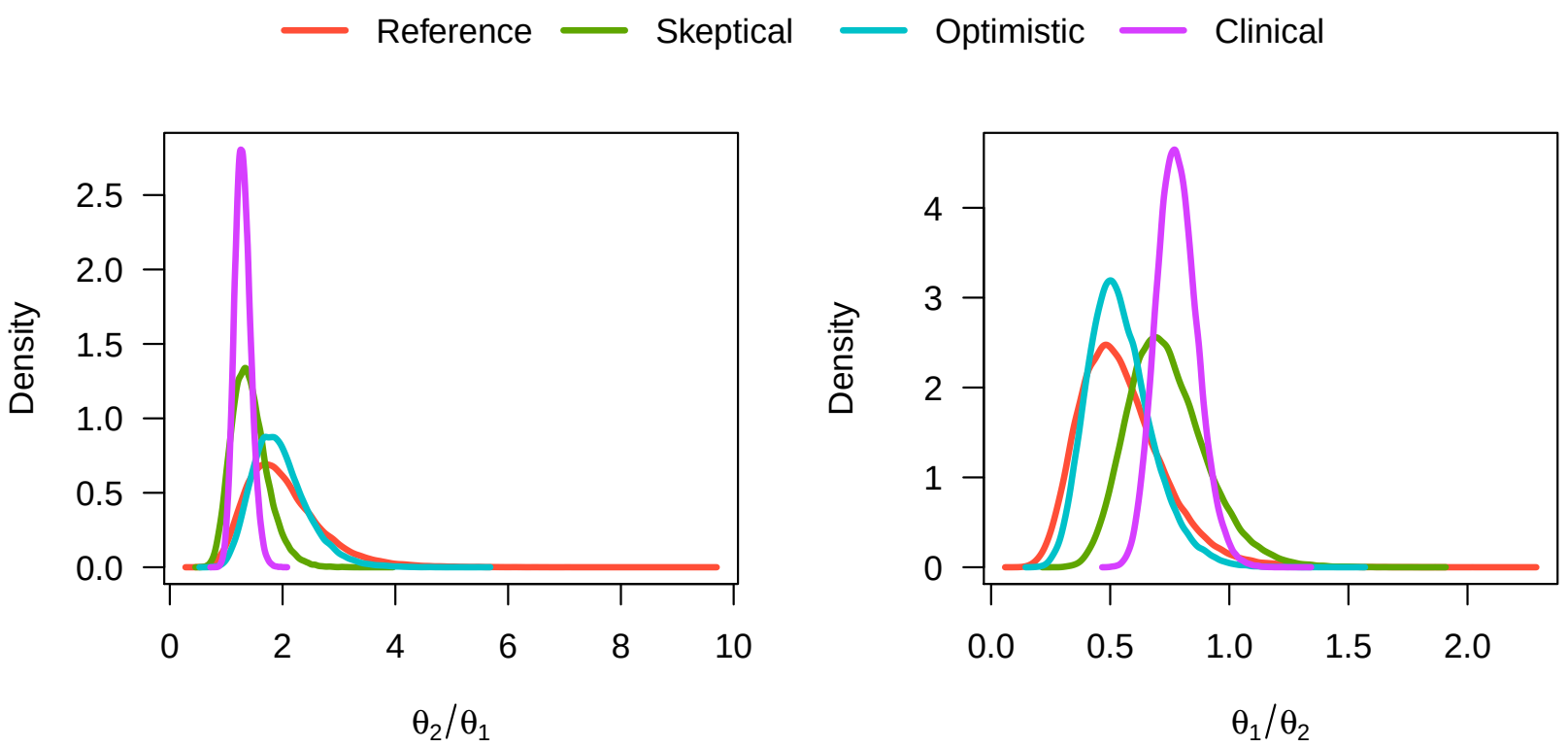
- The bottom line is that this is a fairly small trial (only 36 deaths) and contains insufficient information to substantially alter one's prior beliefs
- This fact is obvious from the Bayesian analysis
- What a person would conclude from a frequentist analysis, however (the p-value, by the way, is 0.05) is anyone's guess

Relative risks

- One final note: I mentioned earlier in the lecture that another reasonable quantity of interest is the relative risk, θ_2/θ_1
- We could certainly analyze this quantity instead of the odds ratio, although the odds ratio is a bit easier to work with when modeling, as $\text{logit}\,\theta$ extends to the whole real line (allowing us to use the normal distribution), unlike $\log\theta$
- However, there is no barrier to parameterizing the model in terms of the odds ratio, then focusing on the posterior for the relative risk

Posterior relative risk

For the GREAT study, the posterior relative risk is similar to the posterior odds ratio:



Case-control designs and relative risks

- It is worth noting, however, that there is an important conceptual difference between the two when it comes to study design
- A common study design in epidemiology is the case-control study, in which instead of tracking individuals to see if they develop a disease or condition (as was done in the GREAT trial), one recruits individuals with (cases) and without (controls) the disease, then determines whether they were exposed to some potentially causative factor in the past
- It is common practice to ignore this fact when analyzing the data, provided that we work with odds ratios instead of relative risks; let's take a quick look at this issue from a Bayesian perspective

A hypothetical study

- Suppose we carried out a case-control study and obtained the following results:

	Disease	No Disease	Total
Exposed	35	15	50
Unexposed	40	60	100
Total	75	75	150

- In reality, a binomial model for the number of exposed subjects is correct – the number of subjects with disease does not actually follow a binomial distribution; but what happens if we analyze the data incorrectly?

Relative risk vs. odds ratio