METRUM INSTITUTE

MI255: Exposure-Response Modeling of Categorical, Count, and Time-to-Event Data Using Bayesian Methods

Bill Gillespie

Metrum Institute

September–December, 2011

Course Introduction

MI255: Exposure-Response Modeling of Categorical, Count, and Time-to-Event Data Using Bayesian Methods provides an introduction to modeling of categorical, count and time-to-event data, and the practical use of WinBUGS for such applications. The course duration and content is equivalent to a single semester 2 credit course at a typical institution of higher learning. Each week's topic will consist of a lecture (one hour) followed by a hands-on lab (one hour). The general plan will be as follows:

- Lectures will be on Mondays at 2 PM EDT.
- Hands-on labs will be on Thursdays at 2 PM EDT (in some cases, the lecture may finish during the first part of the lab on Thursday).

Student Expectations and Requirements for Certificate

- All students are expected to attempt the hands-on exercises prior to the Thursday lab. Instructors will not grade homework assignments, but will review solutions with the entire class on a weekly basis.
- A midterm take-home exam will be assigned at the midpoint of the course.
- A final take-home exam will be assigned at the end of the course (due one week after it is posted).
- Students will be required to complete and submit a modeling project before the end of the course. This project will be based on a real-world (or similar) problem, and will include components of data assembly, model development and evaluation, and a brief report. More details to follow...
- Course grade will be based on the midterm (25%), final (25%) and modeling project (50%)

Course Content Management Site

All students should already have an account to access the main course website. Here's the link: (http://training.metruminstitute.org). Postings to this site will automatically generate an email message to your own e-mail accounts. This site is intended to be the primary repository for all course resources including:

- **News Forum:** Here you'll find updates about class schedule, assignments, etc.
- **Discussion Forum:** Direct your questions about course content to instructors or other students here. You can contribute to ongoing discussions or start a new thread.
- **Technical Support:** Use this forum to submit technical support tickets for problems with any of the Metrum Institute web resources.
- **Link** to the GoToWebinar webcast registration form.
- **Course Materials:** You'll find course notes, examples, and links to recorded lectures under each weekly class heading.

Computer resources

Cloud-based compute server access

- You should have received an email with a username and password for the server.
- Use Remote Desktop Connection to connect to comp2.metruminstitute.org
- Set up a shared folder to exchange files between the server and your computer (to be demo'd).
- R and WinBUGS are installed and your user folder contains the initial course materials.

Computer resources

Use of your own computer

- The initial course materials, including the software we will use throughout the course, are contained in the (rather large) zip file named MI255USB.zip available on the course website.
- For the course we recommend you run the software and course examples directly from a USB flash drive via the shortcuts provided. Install the contents of MI255USB.zip according to:
  1. Download MI255USB.zip to your computer.
  2. Unzip MI255USB.zip. This will create a folder named MI255USB.
  3. Obtain and insert a USB flash drive with $\geq$ 1GB capacity.
  4. Copy the *contents* of MI255USB (not the MI255USB folder itself) to the flash drive.
  5. Check the software installation by double-clicking on the shortcuts named "R 2.13.1.cmd" and "WinBUGS14.cmd". In each case the corresponding program (R or WinBUGS) should be successfully launched. If not, please report it via the Technical Support forum on the course website.

Course Outline

The following course outline was originally developed for a 2 day workshop. All of the listed topics and examples will be covered, but you should anticipate changes to better fit the 1 semester webcast format, e.g., reordering of topics, subdividing tasks within the hands-on examples, and adding new topics and examples.

MI255: Exposure-Response Modeling of Categorical, Count, and Time-to-Event Data Using Bayesian Methods

Objective

Provide an introduction to modeling of categorical, count and time-to-event data, and the practical use of WinBUGS for such applications.

Primary intended audience: pharmacometricians with biological or statistical modeling skills

Background assumed

- PK/PD or statistical modeling
- Some familiarity and hands-on experience with nonlinear regression, mixed effects modeling, Bayesian modeling using WinBUGS and use of R (or S-PLUS).

Course outline

- Some general theory/background:
  - Modeling from a probabilistic point of view: the likelihood function
  - Maximum likelihood for continuous data
  - Extending ML to odd-type data
  - Hierarchical (mixed effects) modeling of odd-type data
  - Bayesian modeling of odd-type data
- Modeling binary data
  - Logistic regression models
  - Bernoulli model for individual binary data
  - Binomial model for summary data
  - Mixed effects modeling of binary data

Course outline

- Hands-on Problem 1: Logistic regression for binary data
- Model evaluation, esp. simulation-based methods for categorical data models
- Hands-on Problem 2: Longitudinal binary data
- Modeling ordered categorical (ordinal) data
  - Cumulative logit models
  - Modeling longitudinal ordinal data: Comparative performance of approximate ML (e.g., NONMEM) and MCMC (e.g., WinBUGS)
- Hands-on Problem 3: Longitudinal ordinal data

Course outline

- Modeling count data
  - The Poisson model
  - Variations on the Poisson model to deal with over-dispersion or zero inflation
- Hands-on Problem 4: Count data
- Modeling time-to-event data for a single event per individual
  - Principles and methods of survival analysis for modeling censored data

Course outline

- Hands-on Problem 5: Time-to-event data: Constant hazard model
- Models with time-varying hazard
- Modeling repeated time-to-event data
  - Modeling of inter-event time intervals
- Hands-on Problem 6: Repeated time-to-event data
- Closing discussion

Some general theory and background

Modeling from a probabilistic point of view

- Those of you who, like myself, were initially trained in a physical or biological science probably started with a deterministic view of modeling.
  - Variability and uncertainty and the statistical tools to deal with them were obligatory nuisances to deal with “noise” but not the focus of the modeling.
  - This began to turn around for many of us with the increasing use of mixed effects modeling in which probability distributions are used to describe the “unexplained” portion of inter-individual variability.
  - Even then I suspect most of us continued to view residual variability as more of a nuisance than as an integral component of the model.
- What we will see today is that such notions of modeling do not translate well to modeling of categorical, count or time-to-event data. A probabilistic perspective is far more useful.

Modeling from a probabilistic point of view: The likelihood function

Start with the notion that the value of a potential future measurement Y is a random variable

- It is not predictable with certainty (even if we know all of the model parameters with certainty)
- The probabilities of different values are described in terms of a probability distribution

$$Y \sim p(y|\theta, x)$$

where θ is a vector of model parameters and x is a vector of covariates.

- If Y is a continuous random variable then $p(y|\theta, x)$ is a probability density function.
- If Y is a discrete random variable then $p(y|\theta, x)$ is a probability function.

Modeling from a probabilistic point of view: The likelihood function

Suppose that future measurement is a plasma drug concentration at some time t following an IV bolus of a drug where the pharmacokinetics can be described by a 1 compartment model with normally distributed residual variation.

$$\begin{aligned} Y &\sim p\left(y|CL, V, \sigma^2, t, D\right) \\ p\left(y|CL, V, \sigma^2, t, D\right) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left(y-\widehat{c}(t,D,CL,V)\right)^2} \\ \widehat{c}(t, D, CL, V) &= \frac{D}{V} e^{-\frac{CL}{V}t} \end{aligned}$$

So even if you know the values of CL, V and σ^2 you cannot say that Y will be a particular value. However, you can say the probability that Y will be within some specified interval.

Modeling from a probabilistic point of view: The likelihood function

- Suppose you already observed a measured value y_{obs}.
- That value is no longer a random variable since we know its value.
- If we insert that observed value into our probability distribution function we now refer to that function as a ***likelihood function***.
- It is the same function as before but we now view it as a ***function of the parameters given the data*** instead of as a function of the data given the parameters.

$$L(\theta|y_{obs}, x) = p(y_{obs}|\theta, x)$$

Modeling from a probabilistic point of view: The likelihood function

$$L(\theta|y_{obs}, x) = p(y_{obs}|\theta, x)$$

- During model development we generally do not know the values of the parameters θ and use the observed data to estimate those parameters.
- The likelihood function contains information about what those parameter values might be.
- We will talk about two different approaches that exploit the likelihood function to estimate θ:
  - Maximum likelihood estimation
  - Bayesian statistical analysis

Maximum likelihood for continuous data

- Apply this idea to our one compartment model example.
- Suppose we observed a plasma drug concentration on two occasions.
- The resulting likelihood function is:

$$\begin{aligned} L\left(CL, V, \sigma^2 | y_{obs1}, y_{obs2}, t_1, t_2, D\right) &= \prod_{i=1}^{2} L\left(CL, V, \sigma^2 | y_{obsi}, t_i, D\right) \\ &= \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left(y_{obsi} - \widehat{c}(t_i, D, CL, V)\right)^2} \\ \widehat{c}\left(t, D, CL, V\right) &= \frac{D}{V} e^{-\frac{CL}{V}t} \end{aligned}$$

Maximum likelihood for continuous data

- The above equation generalizes to any number of observations.
- For n observations described by a normal distribution where the mean is a function $f\left(x, \theta\right)$ of the parameters θ and covariates x:

$$\begin{aligned} L\left(\theta | y_{obs}, x\right) &= \prod_{i=1}^{n} L\left(\theta | y_{obsi}, x_i\right) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left(y_{obsi} - f(x_i, \theta)\right)^2} \end{aligned}$$

- The maximum likelihood estimate of θ is the value of θ that maximizes this likelihood function.

Maximum likelihood for continuous data

- Rather than maximize the likelihood directly, many ML algorithms minimize the transformation $-2log\left(L\left(\theta|y_{obs},x\right)\right)$
- For the above normally-distributed case this becomes:

$$\begin{aligned}
-2\log\left(L\left(\theta|y_{obs},x\right)\right) &= \sum_{i=1}^{n} -2\log\left(L\left(\theta|y_{obsi},x_i\right)\right) \\
&= \sum_{i=1}^{n} -2\log\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}\left(y_{obsi}-f\left(x_i,\theta\right)\right)^2}\right) \\
&= \sum_{i=1}^{n}\left(\log\left(2\pi\right)+\log\left(\sigma^2\right)+\frac{\left(y_{obsi}-f\left(x_i,\theta\right)\right)^2}{\sigma^2}\right) \\
&= n\log\left(2\pi\right)+n\log\left(\sigma^2\right)+\sum_{i=1}^{n}\frac{\left(y_{obsi}-f\left(x_i,\theta\right)\right)^2}{\sigma^2}
\end{aligned}$$

- This shows that the least-squares estimates of θ are also the ML estimates for this case.

Extending ML to “odd-type” data

Binary data

- Most often binary data is used to represent the occurrence or non-occurrence of an event.
- We often use numerical values such as 1 and 0 to represent those two possible outcomes, e.g., 1 for “it happened” and 0 for “it didn’t happen”.
- Suppose we want to model the occurrence of a particular adverse event. Let’s start with one patient. The random variable Y representing the possible AE occurrence is 1 if the AE occurs and 0 if it doesn’t. This is just a Bernoulli trial that is modeled as:

$$\begin{aligned}
Y\sim p\left(y|\theta,x\right) &= \begin{cases} p_{AE}\left(\theta,x\right), & y=1 \\ 1-p_{AE}\left(\theta,x\right), & y=0 \end{cases} \\
&= p_{AE}\left(\theta,x\right)^{y}\left(1-p_{AE}\left(\theta,x\right)\right)^{1-y}
\end{aligned}$$

where $p_{AE}\left(\theta,x\right)$ is the probability that the AE occurs shown as a function of one or more parameters θ and covariates x.

ML modeling of binary data

Now suppose we observe whether or not the AE occurs in 100 patients in a dose response study with the following results:

	Number of patients	
Treatment	Total	AE occurred
Placebo	25	5
10 mg/d	25	7
20 mg/d	25	12
40 mg/d	25	20

ML modeling of binary data

Let's try modeling the probability of an AE in the i^{th} patient as a function of dose according to a linear logistic model:

$$\text{logit}\left(p_{AE}\left(\theta, D_I\right)\right) = \theta_0 + \theta_1 D_i$$

The logit transformation is commonly used to transform between the range of probability (0,1) and the entire real line:

$$\begin{aligned} \text{logit}\left(p\right) &= \log\left(\frac{p}{1-p}\right), \quad 0 < p < 1 \\ \text{logit}^{-1}\left(x\right) &= \frac{e^x}{1+e^x} = \frac{1}{e^{-x}+1}, \quad -\infty < x < \infty \end{aligned}$$

The inverse logit is also sometimes referred to as the expit function.

ML modeling of binary data

So the likelihood for that patient is:

$$L(\theta|y_{obsi}, D_i) = p(y_{obsi}|\theta, D_i) = p_{AE}(\theta, D_i)^{y_{obsi}} (1 - p_{AE}(\theta, D_i))^{1-y_{obsi}}$$

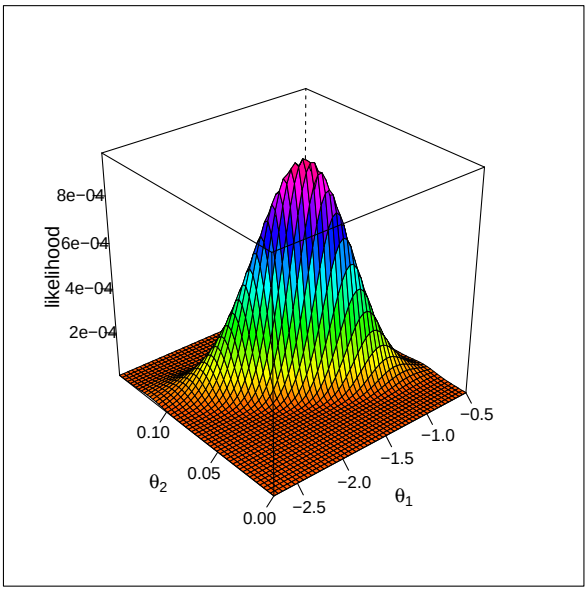
The overall likelihood for the study results is:

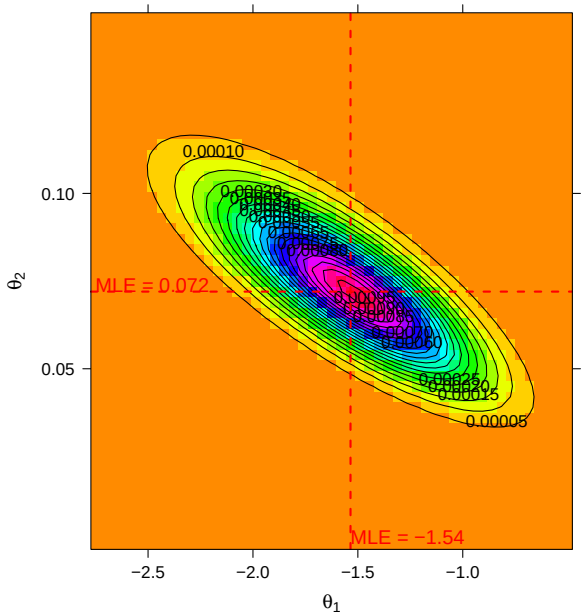
$$\begin{aligned} L(\theta|y_{obs}, D) &= \prod_{i=1}^{100} p(y_{obsi}|\theta, D_i) \\ &= \prod_{i=1}^{100} p_{AE}(\theta, D_i)^{y_{obsi}} (1 - p_{AE}(\theta, D_i))^{1-y_{obsi}} \end{aligned}$$

where yobs and D are vectors of the individual patient values.

The value of θ that maximizes the likelihood value is the ML estimate.

Likelihood function for binary data example





A brief review of Bayesian inference
Bayes Rule

Bayes Rule is the basis for inference about model parameters (θ) given data (y) and prior knowledge about model parameters ($p(\theta)$):

$$\begin{aligned} p(\theta|y) &= \frac{p(\theta)\,p(y|\theta)}{p(y)} = \frac{p(\theta)\,p(y|\theta)}{\int p(\theta)\,p(y|\theta)\,d\theta} \\ &\propto p(\theta)\,p(y|\theta) \end{aligned}$$

The p's are probabilities or probability densities of the specified random variables.

Bayesian modeling/inference process

1. Assess prior distribution $p(\theta)$
   - θ viewed as random variables
   - Subjective
   - Ideally base on all available evidence/knowledge (or belief)
   - Or deliberately select a non-informative (or weakly informative) prior (e.g., reference, vague or improper prior)
2. Construct a model for the data $p(y|\theta)$, also known as the likelihood function when viewed as a function of θ.
3. Calculate posterior distribution $p(\theta|y)$.
   - Use for inferences regarding parameter values
4. Calculate posterior predictive distribution $p(y_{\text{new}}|y)$.
   - Use for inferences regarding future observations

$$p(y_{new}|y) = \int p(y_{new}|\theta)\,p(\theta|y)\,d\theta$$

Bayesian modeling of odd-type data

- Return to the linear logistic regression example where we observe whether or not an AE occurs in each of 100 patients and the probability of an AE is given by

$$\text{logit}\left(p_{AE}\left(\theta, D_I\right)\right) = \theta_1 + \theta_2 D_i$$

- The likelihood function is the same as before.
- Now we must also specify a prior distribution for the model parameters θ. The resulting expression for the posterior distribution of θ is:

$$\begin{aligned} p\left(\theta|y_{obs}, D\right) &\propto p\left(y_{obs}|\theta, D\right) p\left(\theta\right) = L\left(\theta|y_{obs}, D\right) p\left(\theta\right) \\ &\propto \prod_{i=1}^{100} p\left(y_{obsi}|\theta, D_i\right) p\left(\theta\right) \\ &\propto \prod_{i=1}^{100} p_{AE}\left(\theta, D_i\right)^{y_{obsi}} \left(1 - p_{AE}\left(\theta, D_i\right)\right)^{1-y_{obsi}} p\left(\theta\right) \end{aligned}$$

Bayesian modeling of odd-type data

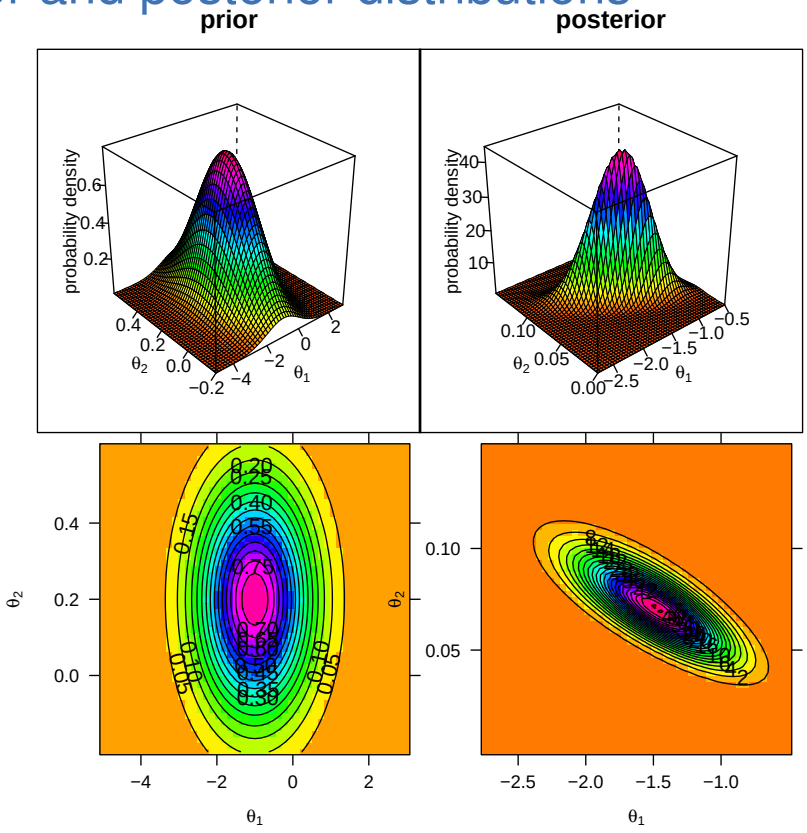
Now suppose we have a little prior information about the value of and choose to represent that knowledge as a bivariate normal distribution with relatively large variances and no correlation, i.e.,

$$\theta \propto N\left(\mu, \Sigma\right)$$

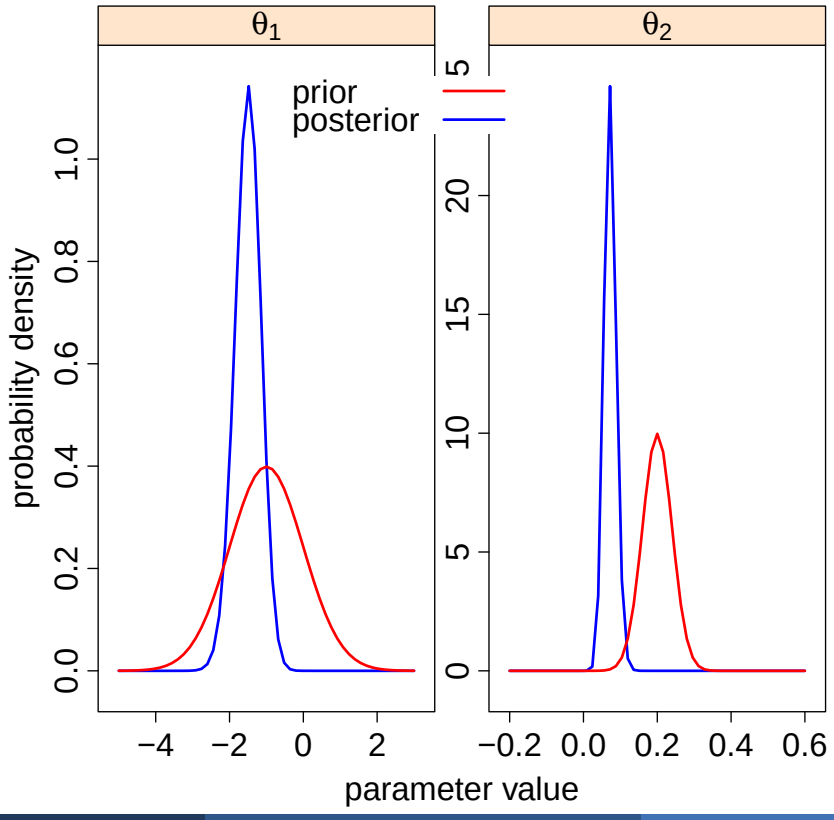
where

$$\mu = (-1, 0.2) \quad \Sigma = \begin{bmatrix} 1^2 & 0 \\ 0 & 0.2^2 \end{bmatrix}$$

Joint prior and posterior distributions



Marginal prior and posterior distributions



Modeling binary data

Modeling binary data: Logistic regression

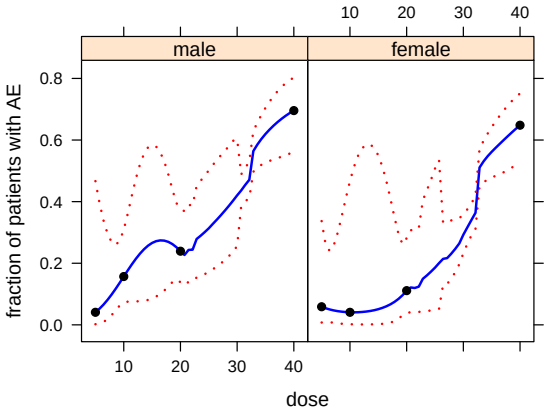
Logistic regression refers to the fitting of binary data with models of the form:

$$\text{logit}\left(p\right) = f\left(x, \theta\right)$$

where p is the probability that some event (e.g., an AE) occurs and x is a vector of covariates.

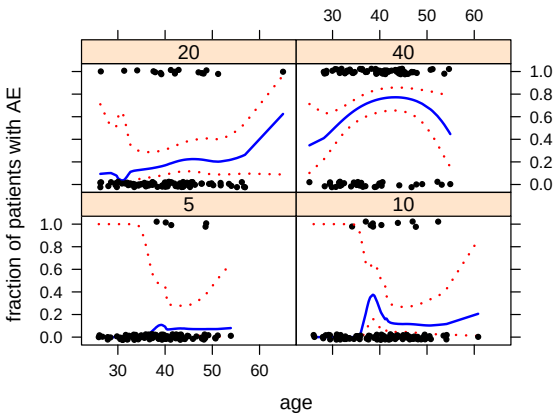
Bernoulli model for individual binary data

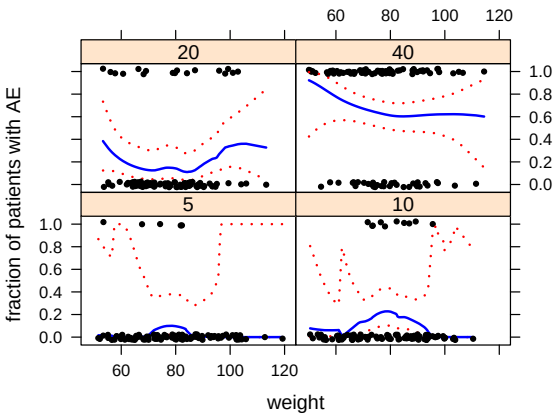
- Suppose we want to model the incidence of a potentially dose-limiting AE as a function of dose in order to support dose selection.
- The data consists of individual patient results from the following study design:
  - Parallel dose-finding study
  - 100 subjects per dose arm
  - Treatment arms: 5, 10, 20 and 40 mg
  - Possible covariates: age, weight, gender



Exploratory analysis indicates:

- Dose-response apparent from summary stats
- Possible effects of gender, weight and age





Proposed model

Linear logistic regression model for AE occurrence as a function of dose, gender, age and weight:

$$\begin{aligned} AE_i &\sim \text{Bernoulli}\left(p_{AE,i}\right) \\ \text{logit}\left(p_{AE,i}\right) &= \theta_1 + \theta_2 D_i + \theta_3 \left(\text{age}_i - 40\right) + \theta_4 \left(\text{weight}_i - 70\right) + \theta_5 \text{gender}_i \end{aligned}$$

Weakly informative priors:

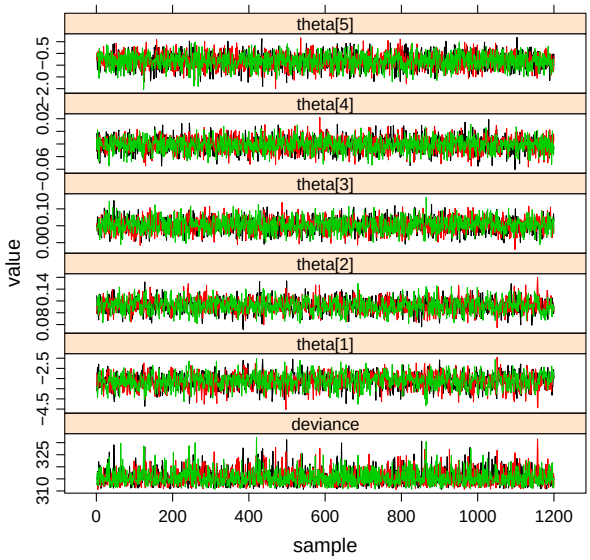
$$\theta_i \sim N\left(0, 10^6\right)$$

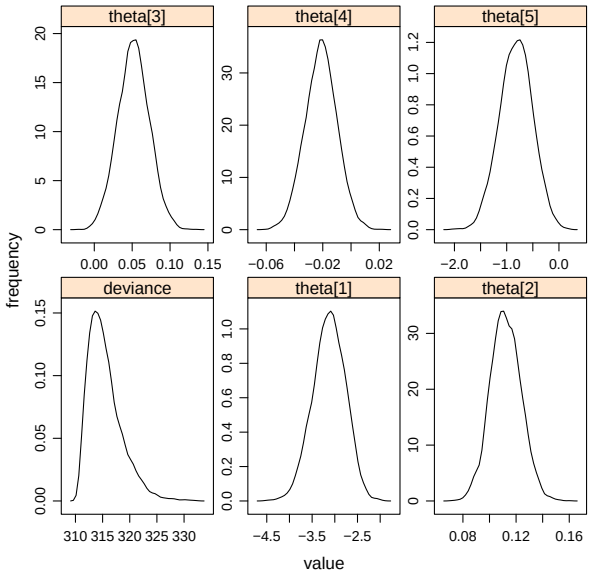
WinBUGS implementation

```
model{

  for(i in 1:nobs){
    ae[i] ~ dbern(p.ae[i]) ## likelihood
    logit(p.ae[i]) <- theta[1] + theta[2]*dose[i] +
      theta[3]*(age[i]-40) + theta[4]*(weight[i]-70) +
        theta[5]*gender[i]

    ae.pred[i] ~ dbern(p.ae[i]) ## posterior predictions
  }

  for(i in 1:5){
    theta[i] ~ dnorm(0,1.0E-6)
  }

}
```

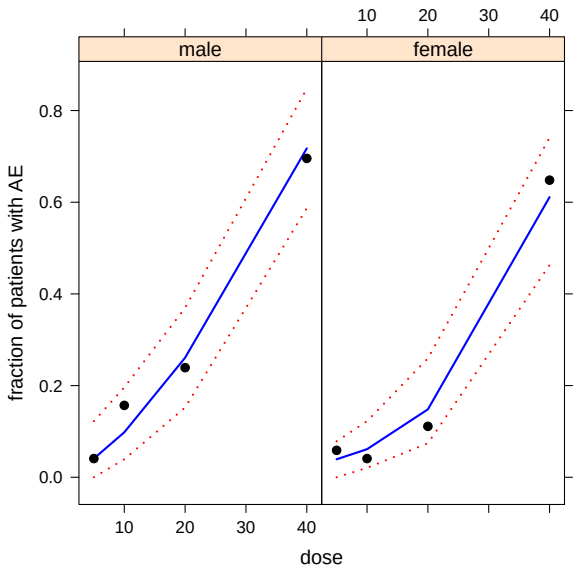
Results

- 10000 $\times$ 3 chains
- burn-in = 4000/chain, thin by 5 (i.e., keep every 5th sample)





- Plot compares observed fractions of patients with the AE to the posterior medians and 90% prediction intervals.
- This is simulated data where the true θ vector is (-2.94, 0.1, 0.05, -0.015, -0.7).



parameter	mean	sd	2.5%ile	25%ile	median	75%ile	97.5%ile	effective N
deviance	315	3.12	311	313	315	317	323	3290
θ_1	−3.13	0.358	−3.84	−3.36	−3.12	−2.88	−2.46	1880
θ_2	0.112	0.0115	0.0896	0.104	0.112	0.12	0.135	1990
θ_3	0.052	0.0206	0.0114	0.038	0.052	0.0656	0.0937	3320
θ_4	−0.0212	0.0113	−0.0435	−0.0286	−0.0211	−0.0136	0.00101	3260
θ_5	−0.821	0.316	−1.45	−1.03	−0.812	−0.605	−0.218	3400

Binomial model for summary data

- A sum of (0,1) Bernoulli distributed (binary) data items is a binomially distributed random variable, i.e., x successes in n trials is equivalent to summing over n (0,1) binary data items where 1 is a "success."
- Probability of a success is modeled in the same manner as for a binary random variable.
- The only difference is in the likelihood function.

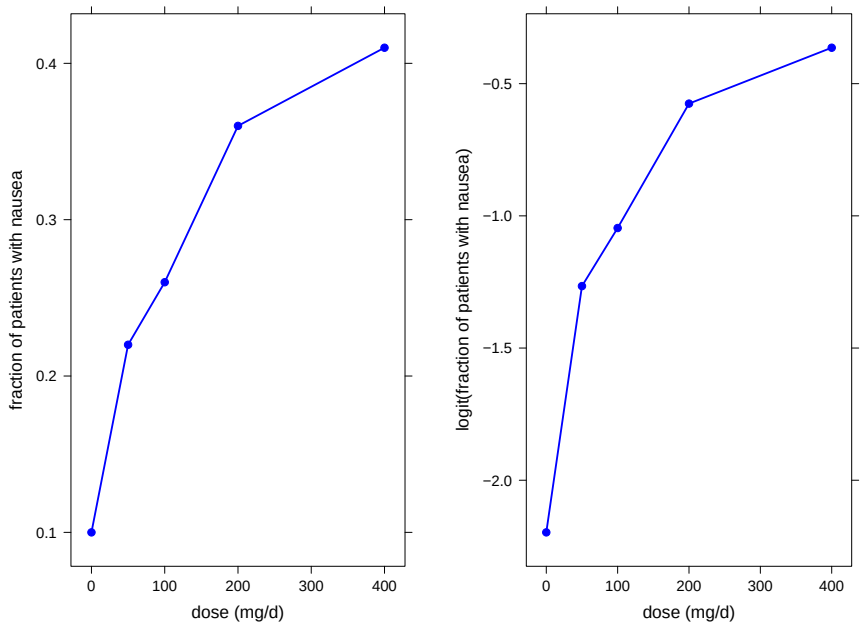
Example: Dose-response model for nausea during treatment with desvenlafaxine

Nausea incidence reported in the Pristiq package insert

dose (mg/d)	number of patients reporting nausea	total number of patients	% reporting nausea
0	64	636	10
50	70	317	22
100	110	424	26
200	111	307	36
400	130	317	41

Example: Dose-response model for nausea during treatment with desvenlafaxine

Nausea incidence reported in the Pristiq package insert



Proposed model

Nonlinear logistic regression model for nausea occurrence in the i^{th} dose group as a function of dose:

$$\begin{aligned} n_{\text{nausea},i} &\sim \text{Binomial}\left(p_{\text{nausea},i}, n_i\right) \\ \text{logit}\left(p_{\text{nausea},i}\right) &= \alpha + \beta D_i^{\gamma} \end{aligned}$$

where $n_{\text{nausea},i}$ is the number of patients reporting nausea and n_i is the total number of patients in the i^{th} dose group, respectively.

Weakly informative priors:

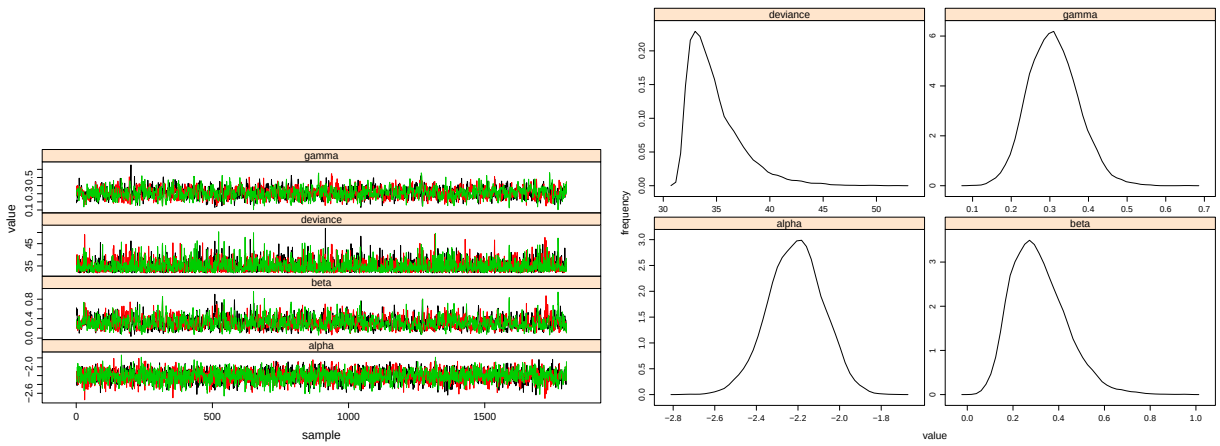
$$\alpha \sim N\left(0, 10^6\right) \quad \beta \sim N\left(0, 10^6\right) \quad \gamma \sim U(0.1, 10)$$

WinBUGS implementation

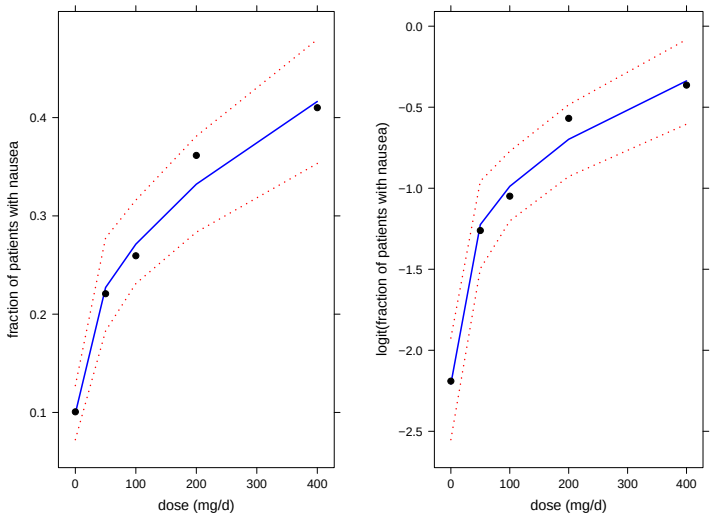
```
model{

  for(i in 1:nobs){
    ## likelihood
    nNausea[i] ~ dbin(pNausea[i],nTotal[i])
    logit(pNausea[i]) <- alpha + beta*pow(dose[i],gamma)

    ## posterior prediction
    nNauseaPred[i] ~ dbin(pNausea[i],nTotal[i])
  }

  alpha ~ dnorm(0,1.0E-6)
  beta ~ dnorm(0,1.0E-6)
  gamma ~ dunif(0.01,10)

}
```


Results

- 100000 × 3 chains
- burn-in = 10000/chain, thin by 50



Plot compares observed fractions of patients reporting nausea to the posterior medians and 90% prediction intervals



parameter	mean	sd	2.5%ile	25%ile	median	75%ile	97.5%ile	effective N
deviance	34.9	2.54	32	33	34.2	36	41.6	4560
α	−2.21	0.131	−2.48	−2.3	−2.21	−2.12	−1.96	3520
β	0.314	0.119	0.13	0.226	0.298	0.386	0.581	2100
γ	0.31	0.0646	0.192	0.264	0.307	0.351	0.445	1740

Hands-on examples: Fictional development program for a mucolytic to treat cystic fibrosis

- Metrum Pharmaceuticals is developing ME-2, a CF treatment that seeks to restore salt transport in the lungs. This is hypothesized to reduce mucous viscosity to near normal thereby increasing mucous clearance from the lungs and improving lung function. The primary efficacy endpoint in confirmatory trials will be the occurrence of pulmonary exacerbations. A pulmonary exacerbation is variously defined in the literature. It generally refers to a deterioration of pulmonary function that requires one or more of a specified set of interventions, e.g., hospitalization, administration of intravenous antibiotics or administration of oral antibiotics.
- In the Metrum trials acute pulmonary exacerbations were defined as an acute exacerbation of CF respiratory symptoms that, in the opinion of the patient's physician, required administration of new oral or intravenous antibiotics.
- Secondary endpoints and biomarkers include pulmonary function measurements such as FEV1 and sputum viscosity. Our hands-on examples will involve the analysis of results from clinical trials conducted in Phases 2 and 3. The order of the examples will not follow the chronological order of the development program.

Hands-on examples: Fictional development program for a mucolytic to treat cystic fibrosis

- Hands-on Problem 1: Logistic regression for binary data
  - Dose-response model for exacerbation occurrence within 24 weeks.
- Hands-on Problem 2: Longitudinal binary data
  - Repeated measures (e.g. each inter-visit period) for AE (e.g., coughing) in Phase II
  - The team hypothesizes that patients will exhibit tolerance for a particular AE, i.e., for a given drug exposure the AE will occur less frequently over time. We will explore this by fitting a longitudinal binary model to the AE data observed in a dose-response study using a model where the probability of an AE can decline with time. This will be used to assess the strength of the evidence supporting the tolerance hypothesis and to explore the effects of dose escalation regimens to reduce AE incidence.

Hands-on examples: Fictional development program for a mucolytic to treat cystic fibrosis

- Hands-on Problem 3: Longitudinal ordinal data
  - Repeated measures QOL score (3 point scale) in Phase 2
- Hands-on Problem 4: Count data
  - Number of coughing episodes in a Phase 2 study
  - Focus on exploring different count models. This is a good setting for hands-on learning about model evaluation and selection.

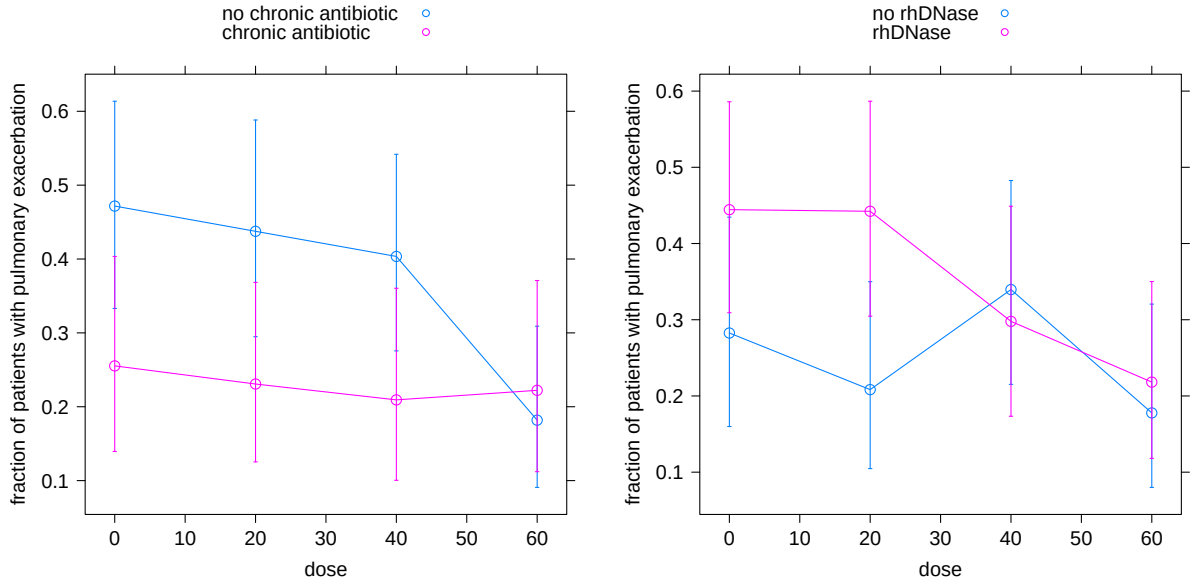
Hands-on examples: Fictional development program for a mucolytic to treat cystic fibrosis

- Hands-on Problem 5: Time-to-event data: Constant hazard model
  - ME-3, a more potent analog of ME-2, is also a mucolytic that is being developed as a treatment for cystic fibrosis. Phase I trials are in progress, so the current focus is on the design of Phase 2 PoC and dose-finding.
  - The development team's preferred primary endpoint for efficacy assessment is time to the first pulmonary exacerbation event, but trials using conventional hypothesis tests require large sample sizes and/or durations in order to achieve adequate statistical power.
  - A model relating FEV1 to exacerbation hazard was developed by meta-analysis of results from past ME-2 trials plus summary data for other mucolytics reported in public sources. It is believed that the model is qualitatively and quantitatively applicable to the new drug candidate because the drugs work by the same mechanism and the patient population is essentially the same.
  - It may be possible to reduce sample sizes or trial duration for a Phase 2 PoC and dose-finding trial by analyzing both exacerbation and FEV1 data using a joint model that incorporates the aforementioned model relating FEV1 to exacerbation hazard—including the informative prior distribution of its parameters. This would permit inferences regarding exacerbations conditioned on prior knowledge and the observed FEV1 and pulmonary exacerbation data.
  - Based on this idea the team conducted a study with fewer patients and half the treatment duration (12 weeks) than a typical study for a CF mucolytic. We analyze the results in this hands-on example.
- Hands-on Problem 6: Repeated time-to-event data
  - Constant hazard model for time between pulmonary exacerbation events

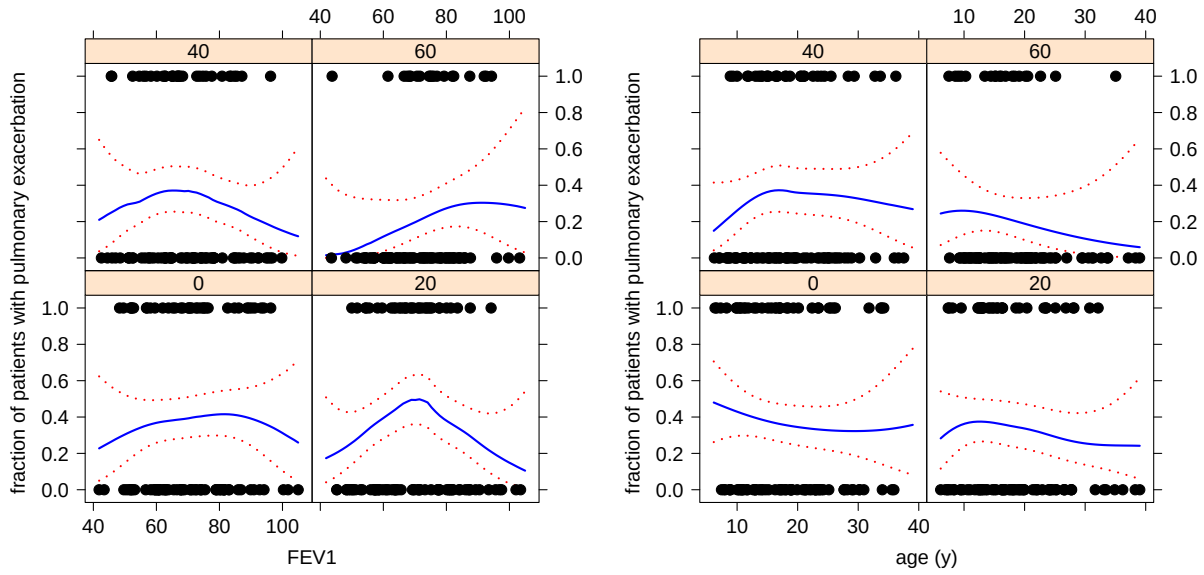
Hands-on problem 1: Logistic regression for binary data

- Phase 2 dose-finding trial in CF patients
  - Parallel design
  - 100 patients per dose arm
  - Multiple doses of ME-2
    - Placebo, 20, 40 and 60 mg qd administered by inhalation for 24 weeks
  - Primary efficacy measurement: Occurrence of ≥ 1 pulmonary exacerbation event within 24 weeks
  - Covariates of possible relevance: age, baseline FEV1, concomitant medications (rhDNase or chronic antibiotic such as azithromycin or inhaled tobramycin)
- Hands-on exercise:
  - Construct a model for occurrence of pulmonary exacerbation as a function of dose and possibly patient-specific covariates.
- Data file: handsOn1/ME2ExacerbationData.csv

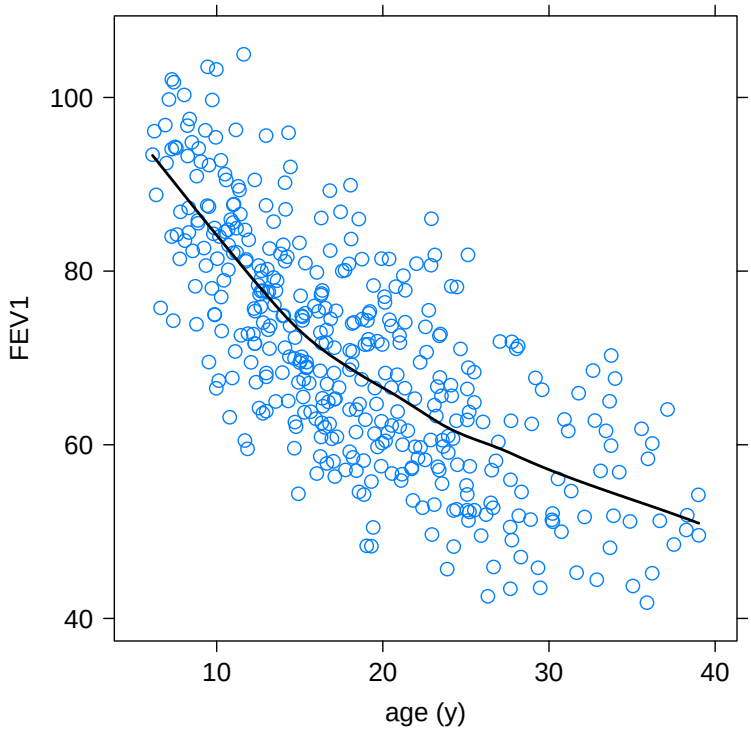
EDA: Fraction of patients with $\geq$ 1 pulmonary exacerbation as a function of dose & selected covariates



EDA: Fraction of patients with $\geq$ 1 pulmonary exacerbation as a function of dose & selected covariates



EDA: Relationship between age and baseline FEV1



Proposed models

Binomial "base" model

- Linear logistic regression model for the number of patients in the i^{th} dose arm with ≥ 1 pulmonary exacerbation as a function of dose:

$$\begin{aligned} n_{exac,i} &\sim \text{Binomial}\left(p_{exac,i}\right) \\ \text{logit}\left(p_{exac,i}\right) &= \theta_i + \theta_2 D_i \end{aligned}$$

- Weakly informative priors: $\theta_i \sim N\left(0, 10^6\right)$

Binary "full" model

- Linear logistic regression model for pulmonary exacerbation occurrence in the i^{th} patient as a function of dose, baseline FEV1 (% of predicted), concomitant medications (chronic antibiotics or rhDNase) and age:

$$\begin{aligned} I_{exac,i} &\sim \text{Bernoulli}\left(p_{exac,i}\right) \\ \text{logit}\left(p_{exac,i}\right) &= \theta_i + \theta_2 D_i + \theta_3\left(\text{FEV1}_i - 70\right) + \theta_4 I_{\text{antibiotic},i} + \\ &\quad \theta_5 I_{\text{DNase},i} + \theta_6\left(\text{age}_i - 18\right) \end{aligned}$$

- Weakly informative priors: $\theta_i \sim N\left(0, 10^6\right)$

Hierarchical (mixed effects) modeling of “odd-type” data

- Suppose that each patient is treated for 4 weeks and reports whether or not the AE occurred each week.
- Now there are 4 binary AE measurements for each patient.
- To account for within patient correlation we will use a mixed effects model that includes inter-patient variation in the model parameters.
- The patient-specific parameter values ϕ_j are modeled according to a parametric probability distribution where the parameters of that distribution (usually the mean) may be functions of patient-specific covariates.
- For our example assume that the inter-patient variation may be described by a normal distribution:

$$\phi_j \sim N(\mu, \Sigma)$$

Hierarchical (mixed effects) modeling of “odd-type” data

The likelihood for an observation in an individual patient y_{obsij} is the same as before except that the parameter values ϕ_j are unique to each patient, i.e.:

$$L\left(\phi_j | y_{obsij}, D_j\right) = p\left(y_{obsij} | \phi_j, D_j\right) = p_{AE}\left(\phi_j, D_j\right)^{y_{obsij}} \left(1 - p_{AE}\left(\theta, D_j\right)\right)^{1-y_{obsij}}$$

where

$$\text{logit}\left(p_{AE}\left(\phi_j, D_j\right)\right) = \theta_{0j} + \theta_{1j} D_j$$

The overall likelihood for the j^{th} patient is:

$$L\left(\phi_j | y_{obs\cdot j}, D_j\right) = \prod_{i=1}^{4} L\left(\phi_j | y_{obsij}, D_j\right)$$

Hierarchical (mixed effects) modeling of "odd-type" data

To make inferences about the population parameters we need to construct the likelihood in terms of $\theta = (\mu, \Sigma)$.

$$
\begin{aligned}
L(\theta|y_{obs}, D) &= \int L(\phi_1, \phi_2, \ldots, \phi_{100}|y_{obs}, D)\, p(\phi_1, \phi_2, \ldots, \phi_{100}|\theta)\, d\phi_1, d\phi_2, \ldots, d\phi_{100} \\
&= \prod_{j=1}^{100} \int L(\phi_j|y_{obs\cdot j}, D_j)\, p(\phi_j|\theta)\, d\phi_j \\
&= \prod_{j=1}^{100} \int \prod_{i=1}^{4} L(\phi_j|y_{obsij}, D_j)\, p(\phi_j|\theta)\, d\phi_j
\end{aligned}
$$

NONMEM approximates this likelihood function using a Laplacian approximation for the integration assuming $p(\phi_j|\theta)$ is a normal distribution. NONMEM 7 now includes additional algorithms that do not use such an approximation.

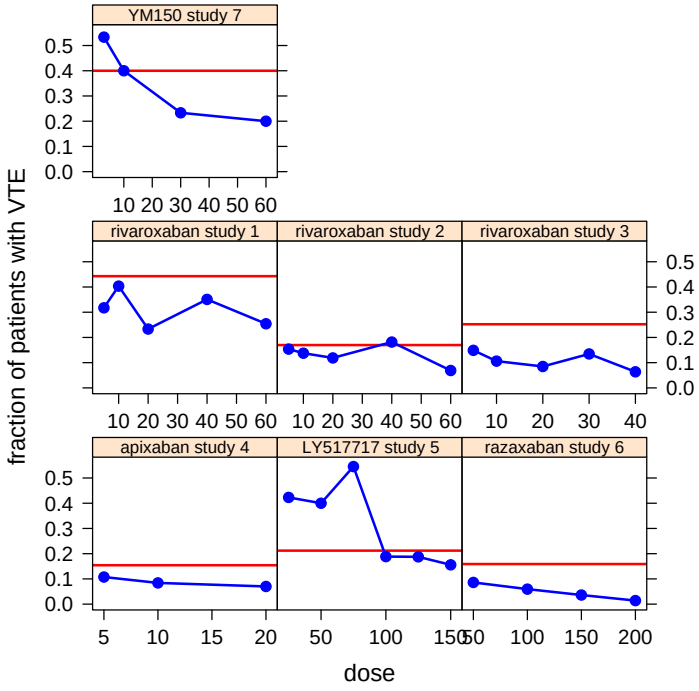
Bayesian treatment of hierarchical (mixed effects) modeling of odd-type data

- Consider again the example where the occurrence of an AE is observed over 4 intervals for each patient.
- In this case we are probably interested in the posterior distributions of the population (θ) and individual parameters (ϕ_j's). The joint posterior distribution of θ and the ϕ_j's is described by:

$$
\begin{aligned}
p(\theta, \phi_1, \phi_2, \ldots, \phi_{100}|y_{obs}, D) &\propto p(y_{obs}|\theta, \phi_1, \phi_2, \ldots, \phi_{100}, D)\, p(\phi_1, \phi_2, \ldots, \phi_{100}|\theta)\, p(\theta) \\
&\propto \prod_{j=1}^{100} p(y_{obs.j}|\phi_j, D_j)\, p(\phi_j|\theta)\, p(\theta) \\
&\propto \prod_{j=1}^{100} \prod_{i=1}^{4} p(y_{obsij}|\phi_j, D_j)\, p(\phi_j|\theta)\, p(\theta)
\end{aligned}
$$

Mixed effects modeling of binary data

- Meta-analysis of factor Xa inhibitor effect on VTE incidence
- Published VTE rates from 7 dose-finding trials comparing new factor Xa inhibitors to enoxaparin. Horizontal lines show observed values for enoxaparin 40-60 mg/d.



Proposed model

- Linear logistic regression model for VTE occurrence in the i^{th} treatment arm of the j^{th} study as a function of dose:

$$\begin{aligned} n_{VTE,ij} &\sim \text{Binomial}\left(p_{VTE,ij}, n_{ij}\right) \\ \text{logit}\left(p_{VTE,ij}\right) &= E_{p,j} + E_{\text{drug},ij} \\ E_{\text{drug},ij} &= \theta_{\text{drug}} \log\left(D_{ij} + 1\right) \\ E_{p,j} &\sim N\left(\theta_p, \sigma^2\right) \end{aligned}$$

  where $n_{VTE,ij}$ is the number of patients experiencing a VTE and n_{ij} is the total number of patients in the in the i^{th} treatment arm of the j^{th} study, respectively.
- Weakly informative priors for drug effects:

$$\theta_{\text{drug}} \sim N\left(0, 10^6\right)$$

- Informative priors for placebo effects (fictional but it illustrates the idea of using historical knowledge about control treatment response):

$$\theta_p \sim N\left(\text{logit}\left(0.45\right), 0.1^2\right) \quad \log\left(\sigma\right) \sim N\left(\log\left(0.4\right), 0.5^2\right)$$

WinBUGS implementation

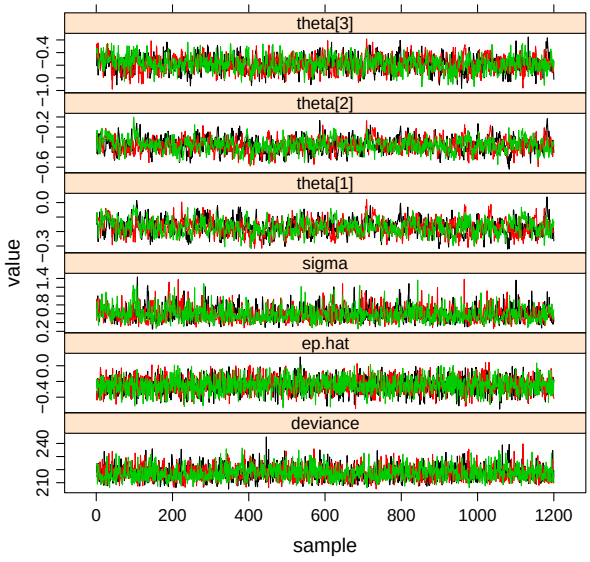
```
model{

  for(i in 1:nobs){
    ## likelihood
    vte[i] ~ dbin(p.vte[i],n[i])
    ## posterior prediction for new observations in same study
    vte.cond[i] ~ dbin(p.vte[i],n[i])
    logit(p.vte[i]) <- ep[study[i]] + edrug[i]
    ## +1 added so that intercept still corresponds to dose = 0
    edrug[i] <- theta[drug[i]]*log(dose[i]+1)

    ## posterior predictions for new observations
    ## in new study of same design
    vte.pred[i] ~ dbin(p.vte.pred[i],n[i])
    logit(p.vte.pred[i]) <- ep.pred[study[i]] + edrug[i]
  }

  for(i in 1:nstudy){
    ## interstudy variation in the intercept
    ep[i] ~ dnorm(ep.hat,tau)
    ep.pred[i] ~ dnorm(ep.hat,tau)
  }
```

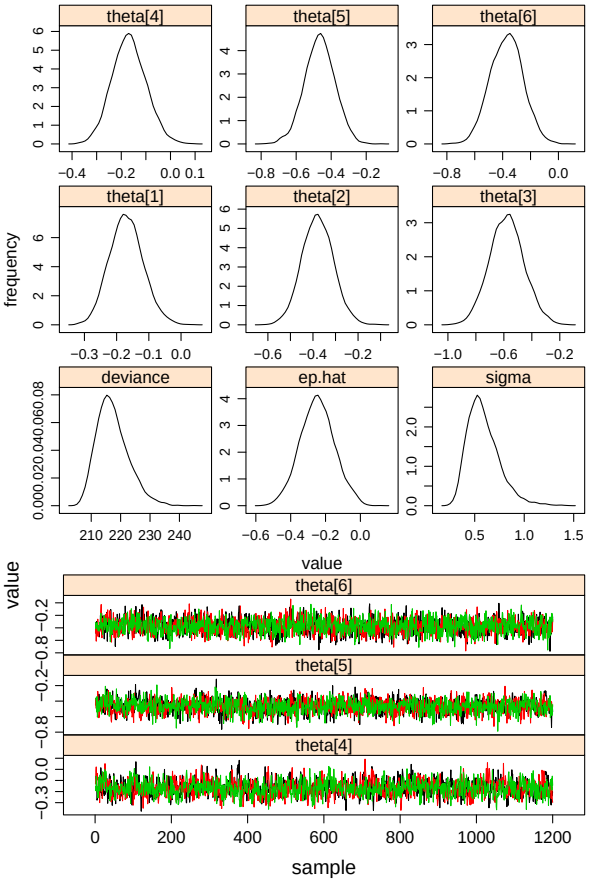

WinBUGS implementation (cont.)

```
  ## informative prior on placebo response
  ep.hat ~ dnorm(ep.hat.prior.mean, ep.hat.prior.precision)
  ep.hat.prior.mean <- logit(0.45)
  ep.hat.prior.precision <- 1/pow(0.1,2)
  log.sigma ~ dnorm(log.sigma.prior.mean, log.sigma.prior.precision)
  log.sigma.prior.mean <- log(0.4)
  log.sigma.prior.precision <- 1/pow(0.5,2)

  log(sigma) <- log.sigma
  tau <- 1/(sigma*sigma)
  for(i in 1:6){
    theta[i] ~ dnorm(0,1.0E-6)
  }

}
```

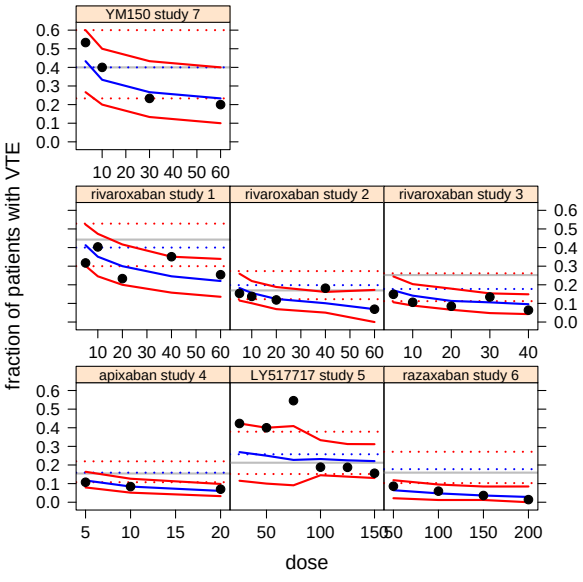
Results

- 10000 × 3 chains
- burn-in = 4000 / chain, thin by 5

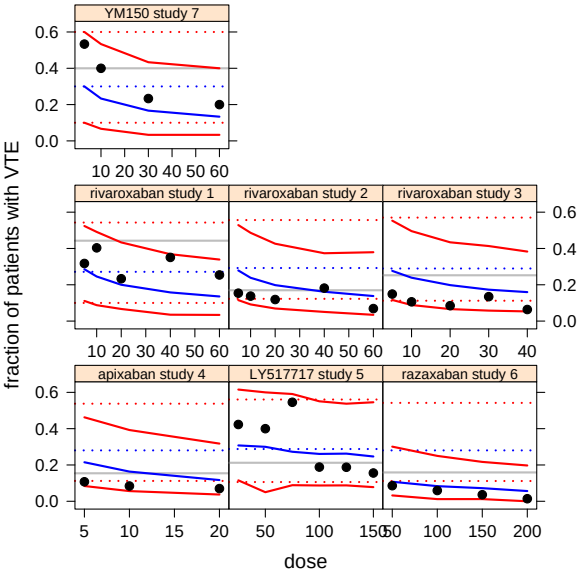




Prediction of new data in same studies



Prediction of new data in new studies of the same design



Horizontal lines show observed (gray) and predicted values for enoxaparin 40-60 mg/d

Model parameter estimates

parameter	mean	sd	2.5%ile	25%ile	median	75%ile	97.5%ile	effective N
deviance	217	5.37	209	214	217	220	230	2420
θ_p	−0.247	0.0976	−0.44	−0.313	−0.248	−0.183	−0.0509	3450
σ	0.59	0.164	0.351	0.472	0.564	0.682	0.984	1710
θ_1	−0.17	0.052	−0.267	−0.205	−0.171	−0.136	−0.0629	708
θ_2	−0.379	0.0693	−0.514	−0.426	−0.38	−0.332	−0.241	809
θ_3	−0.579	0.123	−0.817	−0.662	−0.579	−0.498	−0.331	1130
θ_4	−0.168	0.0689	−0.3	−0.215	−0.17	−0.122	−0.031	1430
θ_5	−0.466	0.0861	−0.64	−0.523	−0.465	−0.409	−0.299	1840
θ_6	−0.371	0.117	−0.597	−0.454	−0.369	−0.293	−0.145	2290

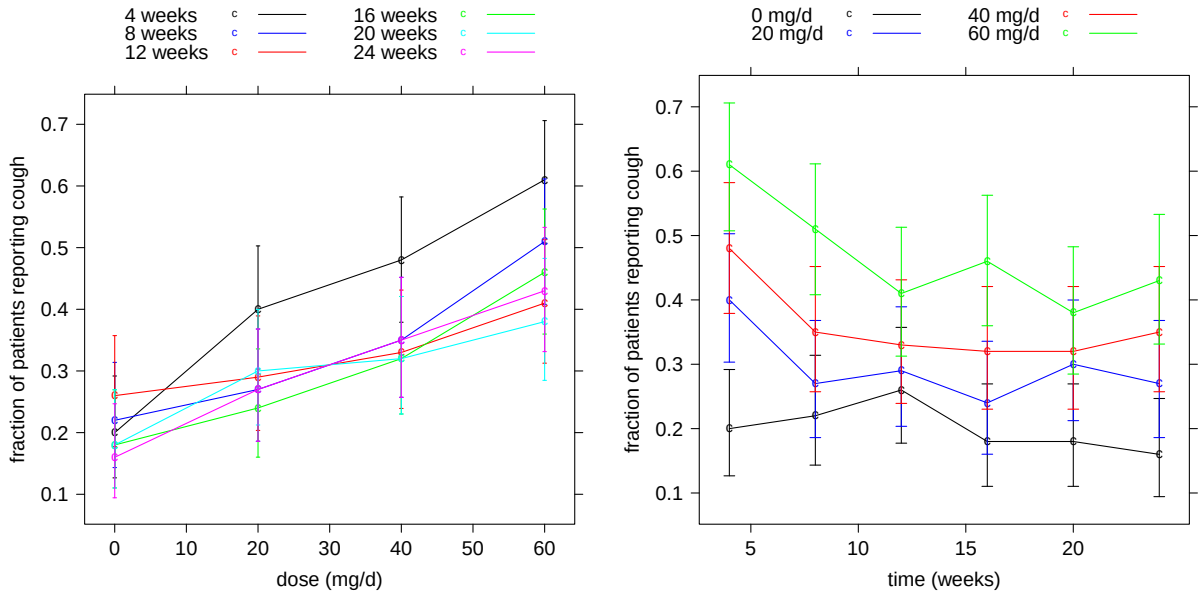
Hands-on Problem 2: Longitudinal binary data

- In the Phase 2 dose-finding trial many patients report moderate to severe coughing.
- Exploratory data analysis suggests that the coughing is drug-related but that it also exhibits tolerance, i.e., the incidence decreases with time.
- We will explore this by fitting a longitudinal binary model to the cough data observed in a dose-response study using a model where the probability of coughing can decline with time.
- This will be used to assess the strength of the evidence supporting the tolerance hypothesis and to explore the effects of dose escalation regimens to reduce AE incidence.

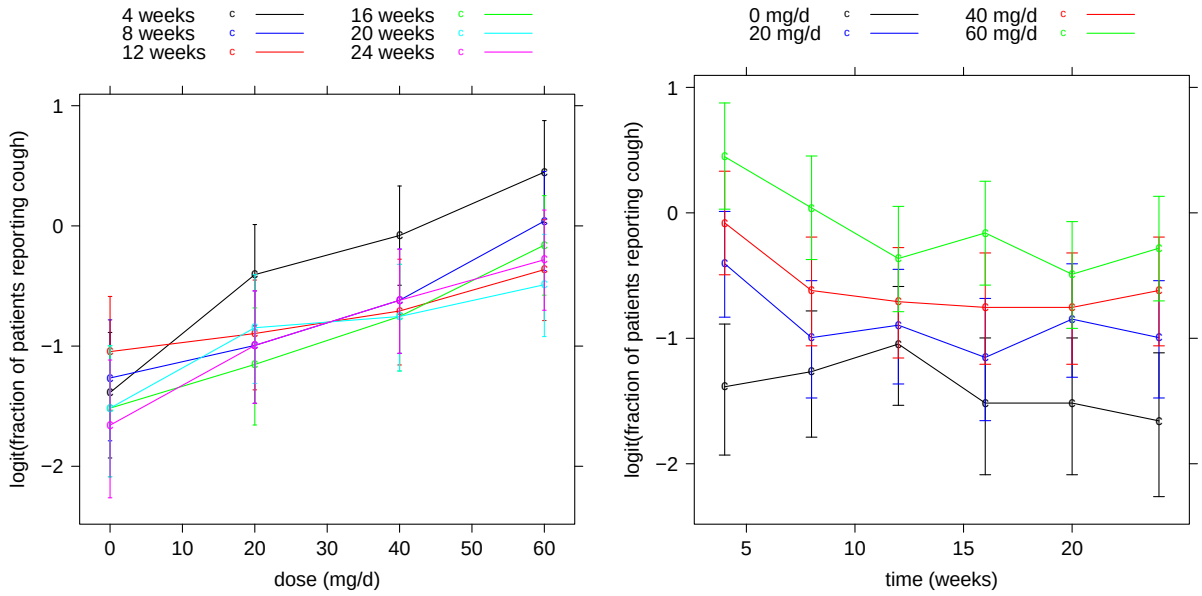
Hands-on Problem 2: Longitudinal binary data

- Phase 2 dose-finding trial in CF patients (same as hands-on problem 1)
  - Parallel design
  - 100 patients per dose arm
  - Multiple doses of ME-2
    - Placebo, 20, 40 and 60 mg qd administered by inhalation for 24 weeks
  - AE measurement: Occurrence of moderate to severe coughing during each 4 week inter-visit period.
- Hands-on exercise:
  - Construct a model for occurrence of coughing as a function of dose and time
- Data file: handsOn2/ME2CoughData.csv

EDA: Fraction of patients reporting moderate to severe coughing as a function of dose and time



EDA: Fraction of patients reporting moderate to severe coughing as a function of dose and time



Proposed model

Logistic regression model where the probability of coughing in each 4 week study period exhibits tolerance described by a model similar to that described by Porchet (JPET 244:231–236 (1988)). Cough occurrence in the j^{th} patient during the i^{th} study period as a function of dose and time:

$$
\begin{aligned}
I_{\text{cough},ij} &\sim \text{Bernoulli}\left(p_{\text{cough},ij}\right) \\
\text{logit}\left(p_{\text{cough},ij}\right) &= E_{0j} + \frac{\alpha D_j}{1 + \frac{x_{tol,ij}}{x_{50}}} \\
x_{tol,ij} &= D_j\left(1 - e^{-k_{tol} t_{ij}}\right) \\
E_{0j} &\sim N\left(\widehat{E}_0, \sigma^2\right)
\end{aligned}
$$

Weakly informative priors:

$$
\begin{aligned}
\widehat{E}_0 &\sim N\left(0, 10^6\right) \quad \alpha \sim N\left(0, 10^6\right) \\
\log\left(x_{50}\right) &\sim N\left(0, 10^6\right) \quad \log\left(k_{tol}\right) \sim N\left(0, 10^6\right) \\
\sigma &\sim U\left(0, 10^4\right)
\end{aligned}
$$

Model evaluation, esp. simulation-based approaches for categorical data models

Bayesian model evaluation & comparison

Typical practical Bayesian model development

- Propose initial model structure based on available prior information.
  - Realistically exploratory analysis of the new data also influences this process,
- Assess whether model is consistent with the data and prior information
  - Posterior predictive checking
  - Are model inferences, predictions and values of parameters or other derived quantities consistent with other knowledge?
- Assess sensitivity to potentially influential assumptions, e.g., choice of prior distributions
- If deficiencies are discovered that could adversely affect important inferences then explore model revisions and reassess as before.

Bayesian model evaluation & comparison

Typical practical Bayesian model development (cont.)

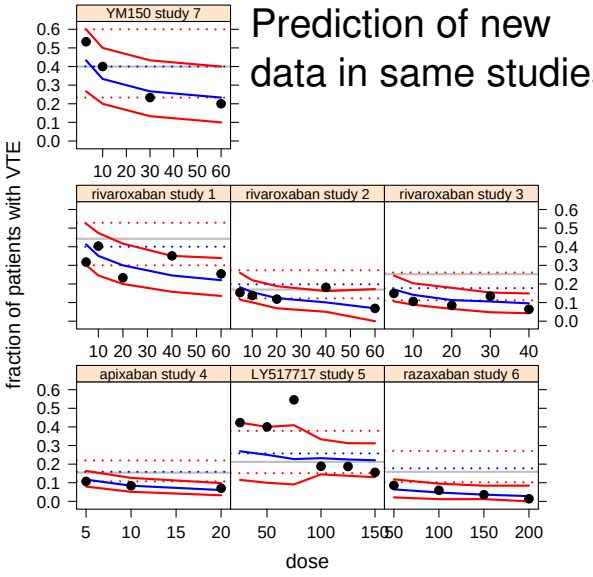
- Use the model resulting from this process for pre-planned inferences.
- Optionally there may be additional hypothesis-generating activities, e.g.,
  - Data mining efforts to explore the influence of covariates not considered in the original model
  - Further exploration of alternative model structures not considered in the previous sensitivity analyses

Posterior predictive checking (PPC)

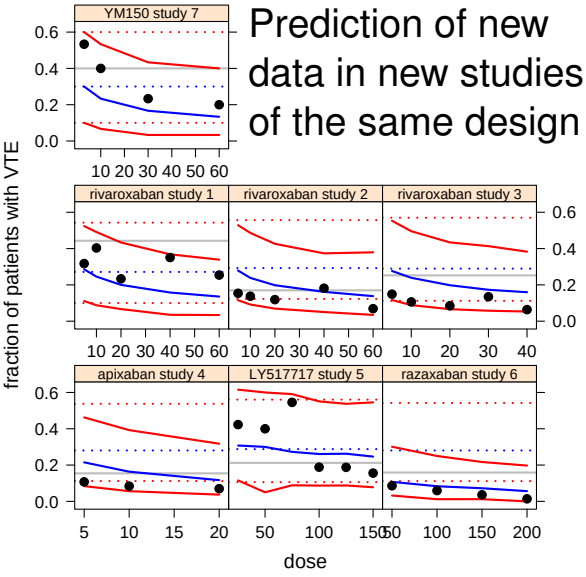
- Graphical checks
  - Comparing data & predictions
  - Comparing data summaries and model predictions/inferences
  - Residuals
- Formal numerical checks
  - Posterior predictive p-values based on a test statistic

Graphical checks: Comparing data & predictions

- Binary data: Plots superimposing posterior predictive distributions and observed data are not very informative.
- Binomial data: Plots of the observed and predicted number or fraction of patients are good model fitting diagnostics.



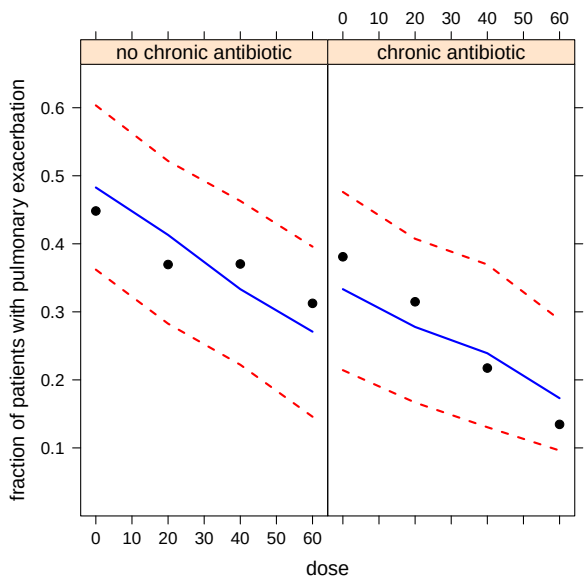
Prediction of new data in same studies

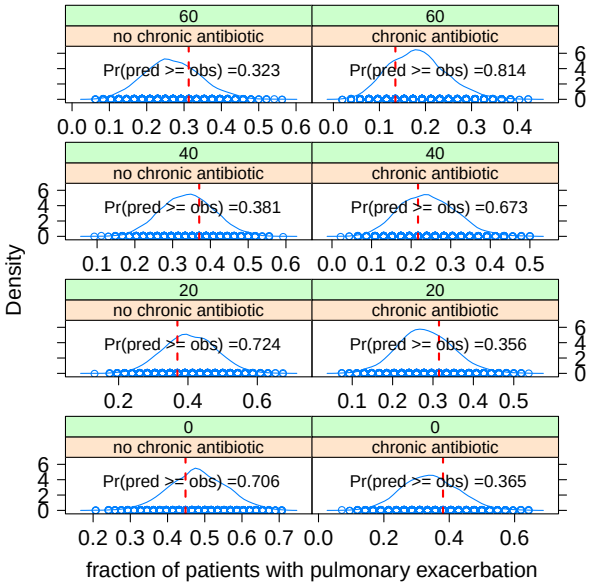


Prediction of new data in new studies of the same design

Graphical checks: Comparing data summaries and model predictions/inferences

Binary data: If the individual binary data can be meaningfully grouped then we can calculate and compare the observed and predicted fraction of patients with a response of 0 or 1 within each group.





Graphical checks: Residuals

- Conventional residuals, $y_i - E(y_i)$, are not as informative or intuitively interpretable for categorical data (particularly in the case of binary data).
- A number of alternatives have been proposed, e.g., normalized, adjusted and deviance residuals.
- In the Bayesian context residuals are distributions that reflect uncertainty in model predictions.

Bayesian residual analysis

- In the binary case residuals are useful for outlier detection (but not much more).
- Let's assume a logistic regression model for binary data of the form:

$$\begin{aligned} y_i &\sim \text{Bernoulli}(p_i) \\ \text{logit}(p_i) &= f(x_i, \theta) \end{aligned}$$

- Albert & Chib proposed 2 types of Bayesian residuals: response residuals and latent response residuals
  - J Albert, S Chib. Bayesian residual analysis for binary response regression models. Biometrika 82(4):747-769 (1995).

Response residual

$$r_i = y_i - p_i$$

- Posterior distributions of the residuals may be plotted.
- Outliers may be detected by identifying residuals with absolute values greater than some specified value.
- But binary response residuals do not have a known sampling distribution, so selection of a critical value based on probability is not easily obtained.

Latent response residual

- The latent variable interpretation of the model may be used to construct a different type of residual.
- The model may be written in the form:

$$\begin{aligned} y_i &= \begin{cases} 1, & z_i > 0 \\ 0, & z_i \leq 0 \end{cases} \\ z_i &= f(x_i, \theta) + \epsilon_i \\ \epsilon_i &\sim \text{Logistic}(0, 1) \end{aligned}$$

- In terms of the latent response z_i we can construct the residual:

$$\epsilon_i = z_i - f(x_i, \theta)$$

Latent response residual

$$\epsilon_i = z_i - f(x_i, \theta)$$

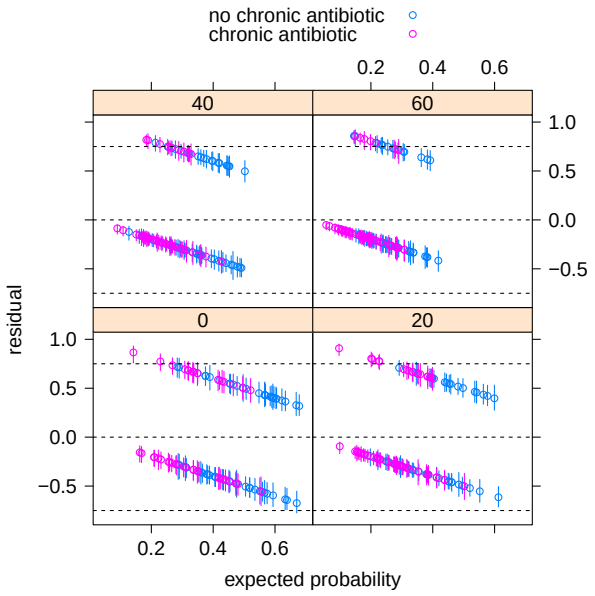
- The posterior density of ϵ_i can be simulated using the posterior distribution of $f(x_i, \theta)$ and the known distribution of ϵ_i given $f(x_i, \theta)$:

$$p(\epsilon_i | f(x_i, \theta)) = \begin{cases} \frac{p_{\text{logistic}}(\epsilon_i)}{\int_{-f(x_i,\theta)}^{\infty} p_{\text{logistic}}(x)dx} I(\epsilon_i > -f(x_i, \theta)), & y_i = 1 \\ \frac{p_{\text{logistic}}(\epsilon_i)}{\int_{0}^{-f(x_i,\theta)} p_{\text{logistic}}(x)dx} I(\epsilon_i < -f(x_i, \theta)), & y_i = 0 \end{cases}$$

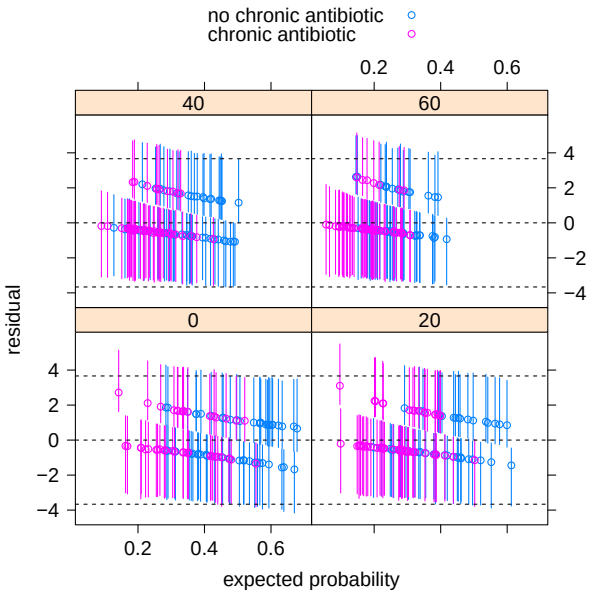
- The resulting latent response residual distribution can be compared to quantiles of the standard logistic distribution.
- This approach has the additional advantage that it is more readily extended to ordinal data.

Bayesian residual examples from Hands-on Problem 1

Response residuals



Latent response residuals



Modeling ordered categorical (ordinal) data

Cumulative logit models

- Suppose we measure some response in terms of an ordinal scale. For example pain measurements might be reported as none, mild, moderate, moderately severe or severe. Often such measurements get reported on a numerical scale such as the integers 1 to 5.
- The basic notion of a cumulative logit model is to convert those ordinal measurements to a collection of binary outcomes
- One general form of a cumulative logit model for an ordinal score from 1 to M is:

$$\begin{aligned} y_i &\sim \text{categorical}\left(p_i|\theta, x_i\right) \\ p_{im} &= \Pr\left(Y = m|\theta, x_i\right) \\ &= \Pr\left(Y \leq m|\theta, x_i\right) - \Pr\left(Y \leq m-1|\theta, x_i\right) \\ \text{logit}\left(\Pr\left(Y \leq m|\theta, x_i\right)\right) &= \alpha_m - f\left(x_i, \theta\right), \quad \alpha_1 < \alpha_2 < \ldots < \alpha_{M-1} \end{aligned}$$

Thus the cumulative probabilities of the ordinal scores share a common model except for the intercept α_m which is unique to each value.

Cumulative logit models

$$
\begin{aligned}
y_i &\sim \text{categorical}\left(p_i|\theta, x_i\right) \\
p_{im} &= \Pr\left(Y = m|\theta, x_i\right) \\
&= \Pr\left(Y \leq m|\theta, x_i\right) - \Pr\left(Y \leq m-1|\theta, x_i\right) \\
\text{logit}\left(\Pr\left(Y \leq m|\theta, x_i\right)\right) &= \alpha_m - f\left(x_i, \theta\right), \quad \alpha_1 < \alpha_2 < \ldots < \alpha_{M-1}
\end{aligned}
$$

- There is one less α than the number of levels.
- Often the α's are parametrized as sums: $\alpha_m = \sum_{i=1}^{m} \Delta\alpha_i$ where $\Delta\alpha_i > 0$ for $i \geq 2$. This is the usual way to enforce the order constraint.

Cumulative logit models

The inequality may also be reversed, i..e.,

$$
\begin{aligned}
y_i &\sim \text{categorical}\left(p_i|\theta, x_i\right) \\
p_{im} &= \Pr\left(Y = m|\theta, x_i\right) \\
&= \Pr\left(Y \geq m|\theta, x_i\right) - \Pr\left(Y \geq m+1|\theta, x_i\right) \\
\text{logit}\left(\Pr\left(Y \geq m|\theta, x_i\right)\right) &= \alpha_m + f\left(x_i, \theta\right), \quad \alpha_2 > \alpha_3 > \ldots > \alpha_M
\end{aligned}
$$

In this case $\alpha_m = \sum_{i=1}^{m} \Delta\alpha_i$ where $\Delta\alpha_i < 0$ for $i \geq 2$.

Cumulative logit models

- The cumulative logit model has the property of proportional odds:

$$\frac{\Pr(Y \leq m|\theta, x_1) / (1 - \Pr(Y \leq m|\theta, x_1))}{\Pr(Y \leq m|\theta, x_2) / (1 - \Pr(Y \leq m|\theta, x_2))} = \frac{e^{-f(x_1,\theta)}}{e^{-f(x_2,\theta)}} = e^{f(x_2,\theta)-f(x_1,\theta)}$$

  In other words the cumulative odds ratio of the same score for 2 different sets of covariate values are independent of the score (m).
- Similarly the cumulative odds ratio of different scores for the same set of covariate values are independent of the covariates:

$$\frac{\Pr(Y \leq m|\theta, x_i) / (1 - \Pr(Y \leq m|\theta, x_i))}{\Pr(Y \leq n|\theta, x_i) / (1 - \Pr(Y \leq n|\theta, x_i))} = \frac{e^{\alpha_m}}{e^{\alpha_n}} = e^{\alpha_m - \alpha_n}$$

Cumulative logit models: Latent variable interpretation

Cumulative logit models may also be interpreted in terms of an underlying continuous regression model. Suppose the observed ordinal variable Y equals m when the unobserved continuous response (a.k.a. latent variable) Z has a value between α_{m-1} and α_m, i.e.,

$$Y = m \iff Z \in [\alpha_{m-1}, \alpha_m)$$

Cumulative logit models: Latent variable interpretation

If Z is distributed according to a logistic distribution with mean $f(x_i, \theta)$ and scale parameter equal to 1:

$$\begin{aligned}
\Pr(Y \leq m|\theta, x_i) &= \Pr(Z \leq \alpha_m|\theta, x_i) \\
&= \int_{-\infty}^{\alpha_m} \frac{e^{-(z-f(x_i,\theta))}}{\left(1+e^{-(z-f(x_i,\theta))}\right)^2} dz \\
&= \frac{1}{1+e^{-(\alpha_m - f(x_i,\theta))}} \\
&= \text{logit}^{-1}(\alpha_m - f(x_i, \theta))
\end{aligned}$$

which is identical to the cumulative logit model described in he previous section.

Modeling longitudinal ordinal data: Comparative performance of approximate ML (e.g., NONMEM) and MCMC (e.g., WinBUGS)

Context & motivation

The AAPS Journal 2004; 6 (3) Article 19 (http://www.aapsj.org).

The Back-Step Method – Method for Obtaining Unbiased Population Parameter Estimates for Ordered Categorical Data

Submitted: October 6, 2003; Accepted: February 20, 2004; Published: August 11, 2004.

Maria C. Kjellsson,[1] Siv Jönsson,[1] and Mats O. Karlsson[1]

[1]Division of Pharmacokinetics and Drug Therapy, Department of Pharmaceutical Biosciences, Uppsala University, Box 591, SE-751 24 Uppsala, Sweden

Journal of Pharmacokinetics and Pharmacodynamics, Vol. 31, No. 4, August 2004 (© 2004)

Estimating Bias in Population Parameters for Some Models for Repeated Measures Ordinal Data using NONMEM and NLMIXED

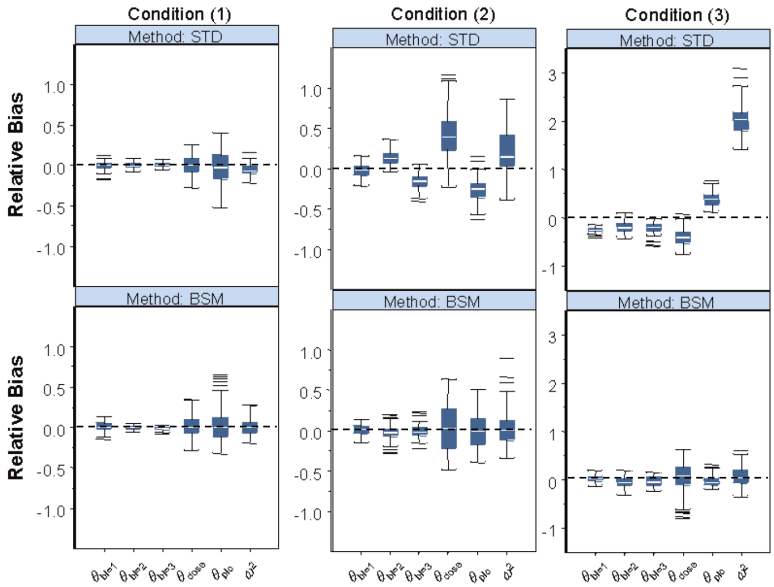
Siv Jönsson,[1,*] Maria C. Kjellsson,[1] and Mats O. Karlsson[1]

Context & motivation

Kjellsson, Jönsson & Karlsson simulation exercises:

- Ordered categorical responses (4 levels)
- NONMEM Laplacian method results in estimation and prediction biases
  - Particularly when the data are skewed to one extreme and/or inter-individual variation (IIV) is large
  - Probabilities of rare events are overestimated
- Illustrated 2 approaches for remedying that bias:
  - The back step method, an iterative application of NONMEM
  - A Gaussian quadrature method (NLMIXED in SAS)

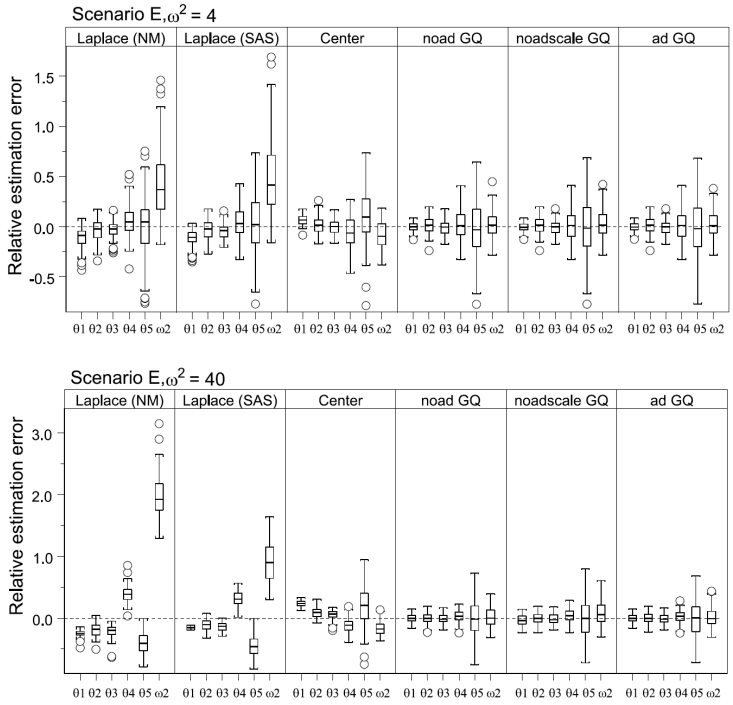
Back step method corrects biases due to the Laplacian approximation



- STD = NONMEM Laplacian method
- BSM = back step method
  - Involves iterative application of the NONMEM Laplacian method

Figure 2 from Kjellsson, Jönsson & Karlsson. The AAPS Journal 2004; 6 (3) Article 19.

Centering helps but some bias remains
Gaussian quadrature shows little or no bias



- SAS NLMIXED used for Gaussian quadrature methods
- Laplacian methods from both NONMEM and SAS were tested
- Center refers to NONMEM Laplacian method with centering

Figure 2 from Jönsson, Kjellsson & Karlsson. J Pharmacokin Pharmacodyn 31(4): 299-320 (2004).

Bayesian modeling using Markov chain Monte Carlo (MCMC) simulation

- Provides results in the form of samples from the joint posterior distribution of the model parameters
- Should not produce the same biases as the Laplacian approximation
- The work presented here tests that expectation by applying MCMC to the same simulated cases as Kjellson et al

Methods

- Trial simulations performed using R
- Same model & parameter values as Kjellsson et al
- Trial design:
  - 4 dose arms: 0, 7.5, 15, 30
  - 250 patients per arm
  - 4 observations per patient (baseline + 3)
  - 100 trial replicates per scenario

Model used for simulation and analysis

The score (0, 1, 2 or 3) at the i^{th} occasion in the j^{th} individual (Y_{ij}) is described by:

$$
\begin{aligned}
Y_{ij} &\sim \text{categorical}\left(p_{ij}|\theta,\omega\right) \\
p_{m,ij} &= \Pr\left(Y_{ij}=m|\theta,\omega,D_j,t_{ij}\right) \\
&= \Pr\left(Y_{ij}\geq m|\theta,\omega,D_j,t_{ij}\right)-\Pr\left(Y_{ij}\geq m+1|\theta,\omega,D_j,t_{ij}\right) \\
\text{logit}\left(\Pr\left(Y_{ij}\geq m|\theta,\omega,D_j,t_{ij}\right)\right) &\sim N\left(\mu_{m,ij},\omega^2\right) \\
\mu_{m,ij} &= \sum_{k=1}^{m}\theta_k+I_{t_{ij}>0}\left(\theta_4+\theta_5 D_j\right), \quad \theta_2<0, \quad \theta_3<0
\end{aligned}
$$

	Parameter values					
Case	θ_1	θ_2	θ_3	θ_4	θ_5	ω^2
1	1.85	-1.85	-1.85	0.483	0.0459	4
2	-4.88	-0.548	-1.18	1.55	0.0303	4
3	-11.8	-1.32	-2.96	3.85	0.717	40

	Expected fraction of baseline scores			
Case	0	1	2	3
1	0.24	0.26	0.26	0.24
2	0.965	0.0122	0.0144	0.0084
3	0.965	0.0122	0.0144	0.0084

NONMEM implementation

```
$PRED

; indicator for post-baseline data
 IPOST = 0
 IF (TIME .GT. 0) IPOST = 1

; treatment effect
 ETREAT = IPOST*(THETA(4) + THETA(5)*DOSE)

; logits for cumulative probabilities
 LPCUM1 = THETA(1) + ETREAT + EXP(THETA(6))*ETA(1) ; SCORE >= 1
 LPCUM2 = LPCUM1 - EXP(THETA(2))                   ; SCORE >= 2
 LPCUM3 = LPCUM2 - EXP(THETA(3))                   ; SCORE >= 3

; cumulative probabilities
 PCUM1 = (1/(1+EXP(-LPCUM1)))
 PCUM2 = (1/(1+EXP(-LPCUM2)))
 PCUM3 = (1/(1+EXP(-LPCUM3)))
```

⋮

NONMEM implementation

⋮

```
; probabilities for each score (likelihood)
 P0 = 1 - PCUM1
 P1 = PCUM1 - PCUM2
 P2 = PCUM2 - PCUM3
 P3 = PCUM3

; indicators for each score
 I0=0
 I1=0
 I2=0
 I3=0
 IF (DV.EQ.0) I0=1
 IF (DV.EQ.1) I1=1
 IF (DV.EQ.2) I2=1
 IF (DV.EQ.3) I3=1

; likelihood
 Y = P0*I0 + P1*I1 + P2*I2 + P3*I3

$ESTIMATION MAX=9999 PRINT=1 METHOD=COND LAPLACE LIKE NOABORT
```

OpenBUGS implementation

Simulated trials analyzed using OpenBUGS + BRugs (R interface to OpenBUGS)

- Model identical to that used for simulation except for presence of prior distributions
- Weakly informative priors
- MCMC settings:
  - 3 chains
  - Burn-in for 4001 samples/chain
  - 5010 post-burn-in samples/chain (keep every 15^{th})

OpenBUGS implementation

```
model{

   for(i in 1:npat){

      ## interpatient variability
      eta[i] ~ dnorm(0,tau.eta)

   }
```

⋮

OpenBUGS implementation

⋮

```
for(i in 1:nobs){

     ## likelihood for observed score
     score[i] ~ dcat(p[i,1:4])

     ## probabilities for each score
     p[i,1] <- 1 - pcum[i,1]
     p[i,2] <- pcum[i,1] - pcum[i,2]
     p[i,3] <- pcum[i,2] - pcum[i,3]
     p[i,4] <- pcum[i,3]

     ## treatment effect model & calculation of cumulative probabilities
     logit(pcum[i,1]) <- theta[1] + (theta[4] +
        theta[5]*dose[i])*(1-equals(time[i],0)) + eta[patient[i]]
     logit(pcum[i,2]) <- logit(pcum[i,1]) + theta[2]
     logit(pcum[i,3]) <- logit(pcum[i,2]) + theta[3]
  }
```

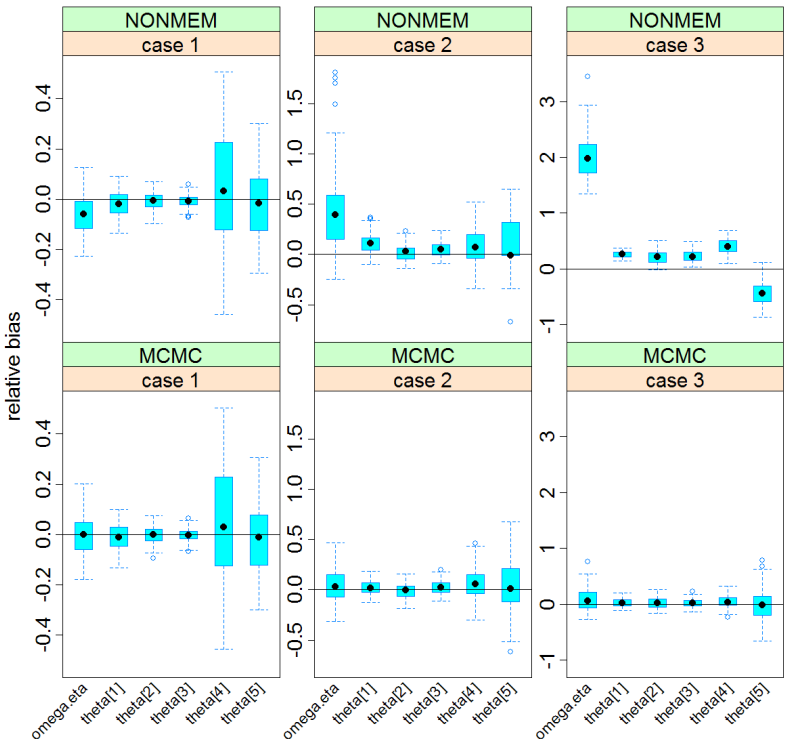
⋮

OpenBUGS implementation

⋮

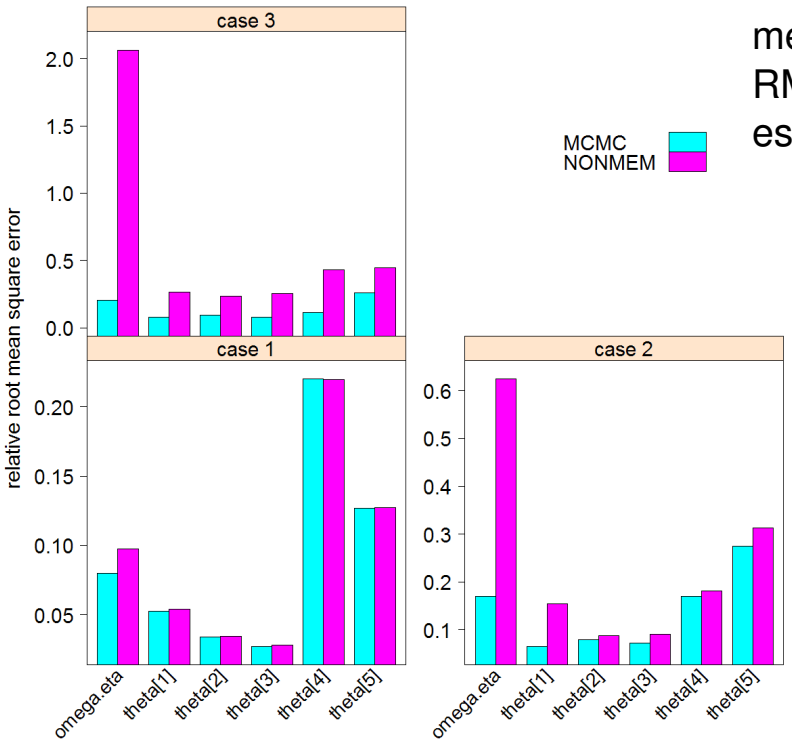
```
  ## prior distributions
  theta[1] ~ dnorm(0,0.00001)
  theta[2] ~ dnorm(-1,0.00001)I(,0)
  theta[3] ~ dnorm(-1,0.00001)I(,0)
  theta[4] ~ dnorm(0,0.00001)
  theta[5] ~ dnorm(0,0.00001)
  sigma.eta ~ dunif(0,1000)
  tau.eta <- 1/(sigma.eta*sigma.eta)

}
```

Relative bias in parameter estimates



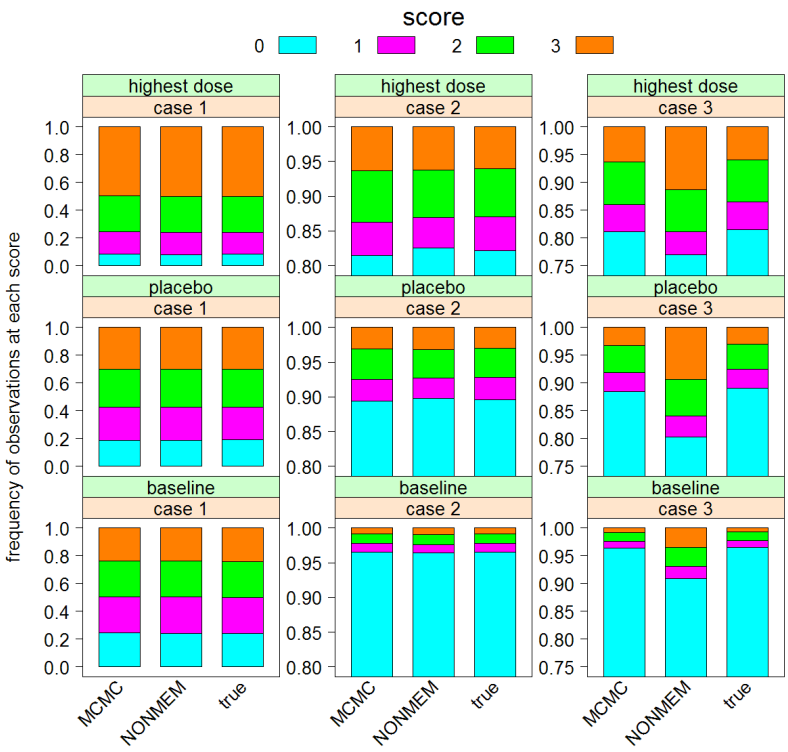
- Bias in NONMEM parameter estimates increases with increasing IIV and skewness
- MCMC estimated posterior means show minimal bias for all 3 cases

Relative root mean square error



MCMC estimated posterior means consistently result in RMSE $\leq$ that for NONMEM estimates

Predicted fraction of responses by score



- When IIV is large biased NONMEM parameter estimates cause overestimation of rare event rates.
- MCMC estimated posterior expected rates show minimal bias

The case for Bayesian modeling of ordinal data using MCMC

- Better estimation and prediction performance than methods using linear or Laplacian approximation to the likelihood
- Yields an estimate of the entire joint posterior distribution of the model parameters
  - Describes uncertainty in parameters
  - Uncertainty in derived quantities, e.g., predictions, is easily calculated from MCMC samples
- Can easily and rigorously include prior information
- Available tools, e.g., WinBUGS/OpenBUGS, permit very flexible model specification:
  - Rich collection of built-in probability distributions
  - No limit on levels of variability

The case against Bayesian modeling of ordinal data using MCMC

Requires more computation time

- $\sim$ 15–45 minutes per trial (elapsed time with Intel Core Duo 2.33 GHz, 2 GB RAM)
- Limited benefit from parallel computation
  - Though substantial gains are possible by running multiple chains in parallel
- NONMEM requires substantially less time to obtain point estimates
- SAS NLMIXED using Gaussian quadrature is also faster
  - But if you want rigorous characterization of uncertainty with ML methods:
    - Bootstrapping is probably the best option
    - And that also requires sizable computation time
    - But it is readily accelerated via parallel computation

Generalizations of the cumulative logit model: Other cumulative link models

Models for ordinal data may be constructed using link functions other than the logit function:

$$g\left(\Pr\left(Y_i \leq m|\theta, x_i\right)\right) = \alpha_m - f\left(x_i, \theta\right), \quad \alpha_1 < \alpha_2 < \ldots < \alpha_{M-1}$$

g can be the inverse of any cumulative distribution function (cdf). A couple more commonly used link functions are:

- Cumulative probit models where g is the inverse of the standard normal cdf:
$$g^{-1}\left(x\right) = \Phi\left(x\right) = \int_{-\infty}^{x} \frac{1}{2\pi} e^{u^2/2} du$$
- Complementary log-log models where g is the inverse cdf of the extreme value distribution:
$$g\left(x\right) = \log\left(-\log\left(1 - x\right)\right)$$

Generalizations of the cumulative logit model: Other latent variable models

- The latent variable interpretation of cumulative logit models suggests a whole range of models that could be constructed based on continuous latent variables with distributions other than the logistic distribution with a constant scale parameter.
- That is a conceptually attractive approach when there is a mechanistic rationale, e.g., when there is a causal relationship between some well-understood but unobserved continuous response and the observed ordinal value.

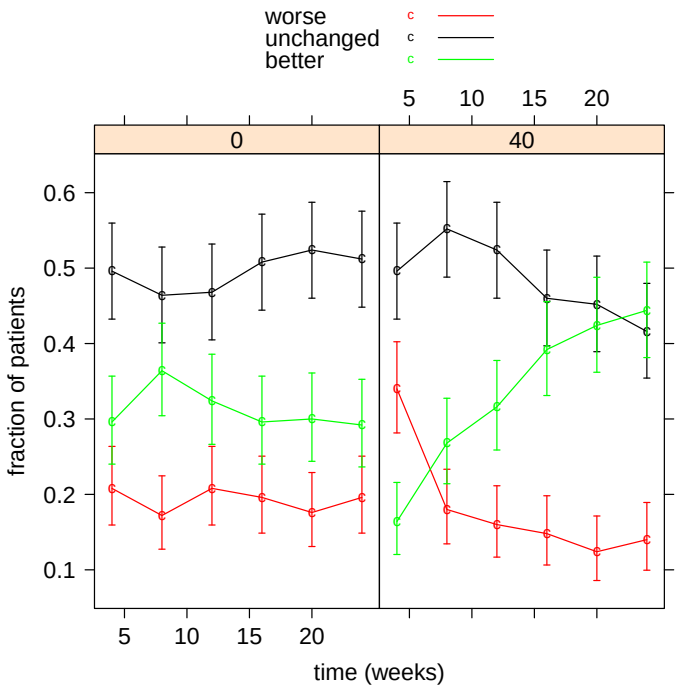
Hands-on problem 3: Longitudinal ordinal data

- At each visit during a Phase 3 confirmatory trial patients indicated whether their health-related quality of life (HRQOL) was worse, unchanged or better than it was prior to the trial.
- Thus HRQOL is a 3 level ordinal scale.
- Exploratory data analysis suggests that HRQOL initially worsens for those patients assigned to ME-2 compared to those on placebo (perhaps due to the coughing), but improves over time.
- We will explore this by fitting a longitudinal cumulative logit model to the HRQOL data observed in the Phase 3 study.

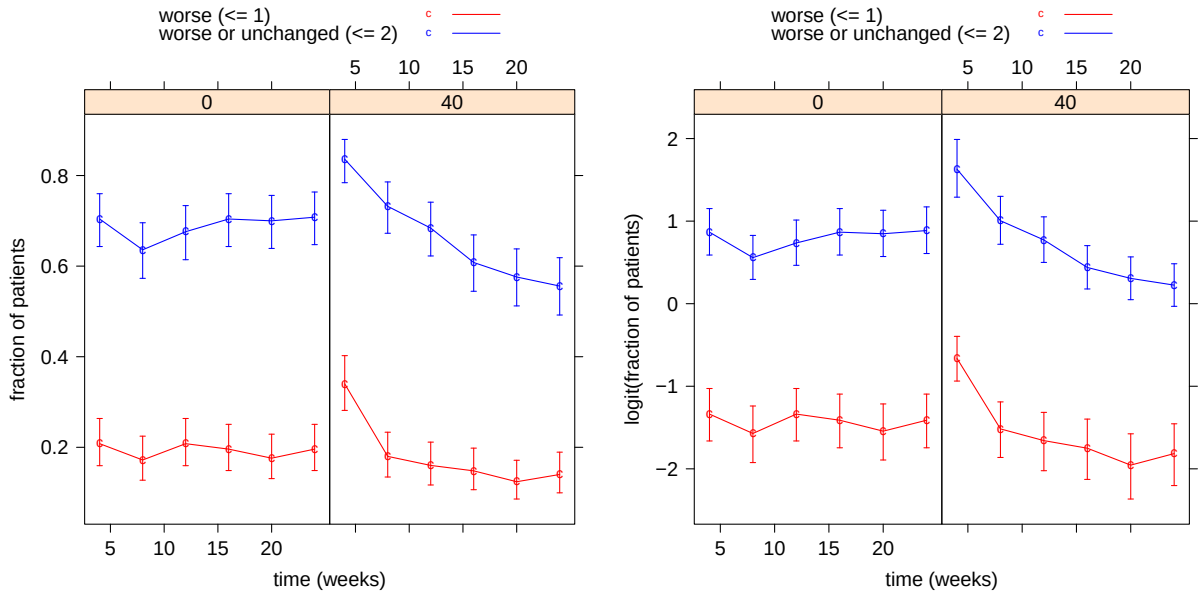
Hands-on problem 3: Longitudinal ordinal data

- Phase 3 confirmatory trial in CF patients
  - Parallel design
  - 250 patients per treatment arm
  - Multiple doses of ME-2
  - Placebo vs. 40 mg qd administered by inhalation for 24 weeks
  - HRQOL measurement: 3 level ordinal score (worse, unchanged, better) reported every 4 weeks.
- Hands-on exercise:
  - Construct a model for the HRQOL score as a function of dose and time.
- Data file: handsOn3/ME2QOLData.csv

EDA: Fraction of patients reporting each level of the HRQOL score as a function of dose and time



EDA: Fraction of patients reporting an HRQOL score $\leq$ 1 or 2 as a function of dose and time



Proposed model

Cumulative logit model for the HRQOL score on the i^{th} visit of the j^{th} patient:

$$
\begin{aligned}
QOL_{ij} &\sim \text{categorical}\left(p_{ij}\right) \\
p_{m,ij} &= \Pr\left(QOL_{ij} = m\right), \quad m \in \{1,2,3\} \\
&= \Pr\left(QOL_{ij} \leq m\right) - \Pr\left(QOL_{ij} \leq m-1\right) \\
\text{logit}\left(\Pr\left(QOL_{ij} \leq m\right)\right) &= \alpha_m - \left(E_{\text{placebo},ij} + E_{\text{drug},ij}\right) + \eta_j \\
E_{\text{placebo},ij} &= \beta t_{ij} \\
E_{\text{drug},ij} &= I_{D_j>0}\left(a + b\left(1 - e^{-kt_{ij}}\right)\right) \\
\eta_j &\sim N\left(0, \sigma^2\right)
\end{aligned}
$$

Weakly informative priors:

$$
\begin{aligned}
\alpha_1 &\sim N\left(0, 10^6\right) \quad \alpha_2 - \alpha_1 \sim \text{truncated } N\left(0, 10^6\right), \ \alpha_2 - \alpha_1 > 0 \\
\beta &\sim N\left(0, 10^6\right) \quad a \sim N\left(0, 10^6\right) \quad b \sim N\left(0, 10^6\right) \\
\log\left(k\right) &\sim N\left(0, 10^6\right) \quad \sigma \sim U\left(0, 10^4\right)
\end{aligned}
$$

Modeling count data

Modeling count data

- “Count data” (number of times an event occurs within a specified time interval) is another type of discrete random variable.
- Unlike ordinal data it does not have a fixed upper bound.
- Event counts may be conceptualized as observed manifestations of the underlying hazard for event occurrence.
- A hazard is (loosely speaking) the instantaneous probability density for event occurrence (more on this when we get to time-to-event data) or, somewhat more intuitively, it is the expected instantaneous event rate.

Modeling count data: Examples

Vomiting events: E H Cox, C Veyrat-Follet, S L Beal, E Fuseau, S Kenkare, L B Sheiner. A Population Pharmacokinetic-Pharmacodynamic Analysis of Repeated Measures Time-to-Event Pharmacodynamic Responses: The Antiemetic Effect of Ondansetron. J Pharmacokin Biopharm 27:625–644 (1999).

Seizures: R Miller, B Frame, B Corrigan, P Burger, H Bockbrader, E Garofalo, R Lalonde. Exposure-response analysis of pregabalin add-on treatment of patients with refractory partial seizures. Clin Pharmacol Ther 73:491–505 (2003).

Urinary incontinence events: S K Gupta, G Sathyan, E A Lindemulder, P-L Ho, L B Sheiner, L Aarons. Quantitative characterization of therapeutic index: Application of mixed-effects modeling to evaluate oxybutynin dose-efficacy and dose-side effect relationships. Clin Pharmacol Ther 65:672–684 (1999).

Neonatal apneic events: C J Godfrey. Mixed effects modeling analysis of count data. In E I Ette, P J Williams. Pharmacometrics: The Science of Quantitative Pharmacology, Wiley, 2007, pp 699–721,

The Poisson model

Standard (constant hazard) model

- The Poisson process
  - If the following conditions are true:

    Constant hazard: the probability of an event within any very short time interval is (approximately) proportional to the length of that interval.

    Independence: the numbers of events in any 2 disjoint intervals are independent.

    Discrete events: only one event can occur at any instant of time.
  - then the number of events in any time interval (t_{i-1}, t_i) is a random variable distributed according to:

$$\Pr\left(N\left(t_i\right) - N\left(t_{i-1}\right) = n\right) = \frac{1}{n!}\left(h\left(t_i - t_{i-1}\right)\right)^n e^{-h\left(t_i - t_{i-1}\right)}$$

    where $N(t)$ is the cumulative number of events from time = 0 to time = t.
  - In other words:

$$N\left(t_i\right) - N\left(t_{i-1}\right) \sim \text{Poisson}\left(h\left(t_i - t_{i-1}\right)\right)$$

Properties of the Poisson distribution

$$
\begin{aligned}
x &\sim \text{Poisson}(\lambda), \quad x = 1, 2, \ldots, \infty \\
p(x) &= \frac{1}{x!}\lambda^x e^{-\lambda} \\
E(x) &= \lambda \\
\text{Var}(x) &= \lambda
\end{aligned}
$$

If the random variables x_i, $i = 1, 2, \ldots, n$ are independent and

$$x_i \sim \text{Poisson}(\lambda_i)$$

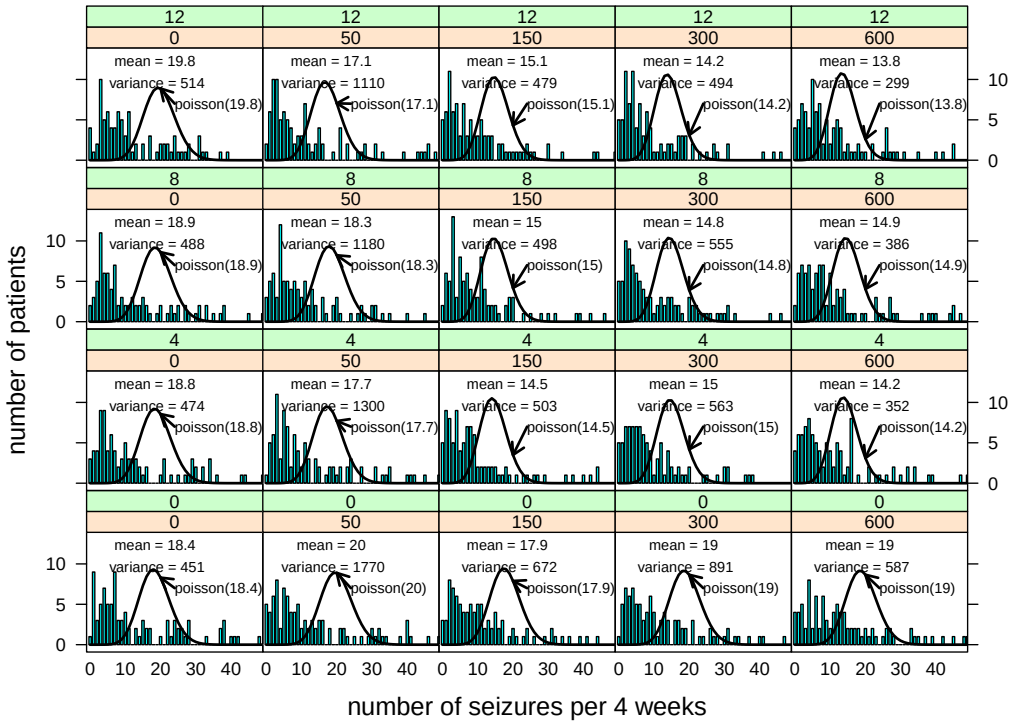
then

$$\sum_{i=1}^{n} x_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right)$$

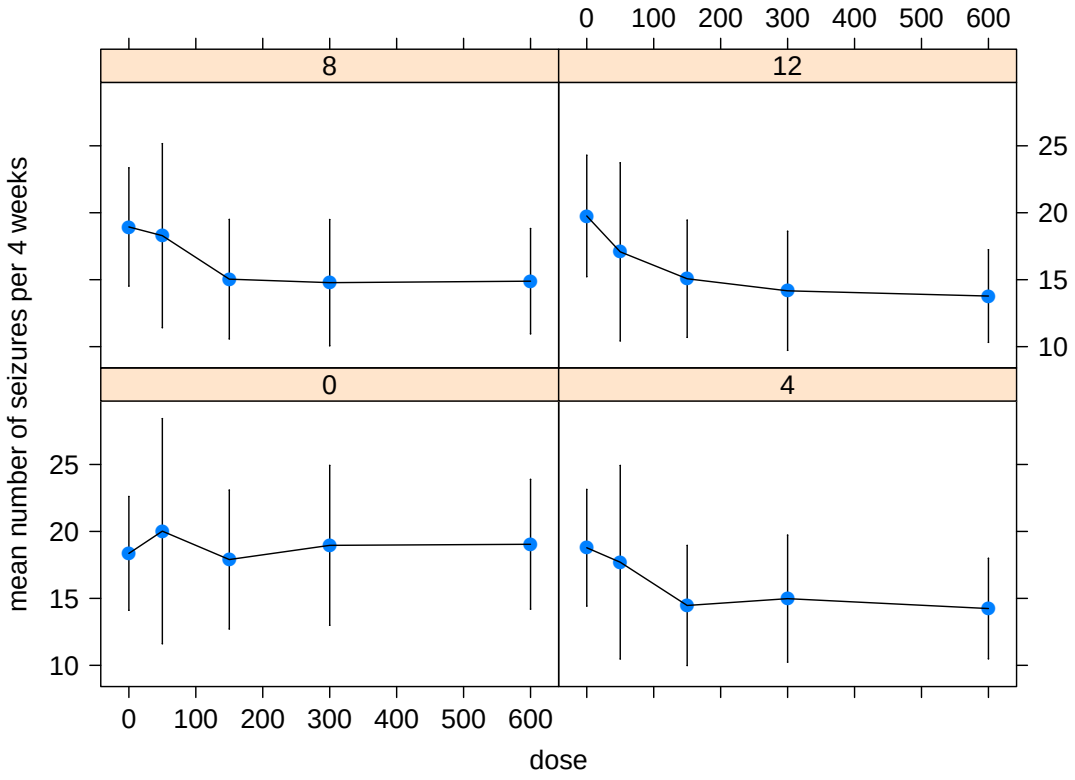
Count data example: Effect of pregabalin on seizure frequency

- Simulated data based on model reported by Miller et al (Model 3).
- Pregabalin 0, 50, 150, 300 and 600 mg/d
- 100 patients per dose arm
- 4 week baseline phase
- 12 week double blind phase
- Data consists of number of seizures in each 4 week period

Empirical distributions of seizure counts by dose and time compared to Poisson distributions with the same means



Mean seizure count vs. dose by week



Mixed effects model for the longitudinal seizure count data

Bayesian implementation of Miller et al Model 3 for number of seizures experienced by the j^{th} patient during the i^{th} period:

$$
\begin{aligned}
n_{\text{seizure},ij} &\sim \text{Poisson}\left(\lambda_j\right) \\
\lambda_j &= \begin{cases} \text{base}\, e^{\eta_j}, & \text{baseline period} \\ \text{base}\left(1 - E_{\text{placebo}} - E_{\text{drug},j}\right) e^{\eta_j}, & \text{otherwise} \end{cases} \\
E_{\text{drug},j} &= \frac{E_{\max} D_j}{ED_{50} + D_j} \\
\eta_j &\sim N\left(0, \omega^2\right)
\end{aligned}
$$

Weakly informative priors including transformation to constrain $E_{\max} + E_{\text{placebo}} < 1$ and $E_{\text{placebo}} < 1$:

$$
\begin{aligned}
\log\left(p_1\right) &\sim N\left(0, 10^6\right) \quad \log\left(p_2\right) \sim N\left(0, 10^6\right) \\
E_{\text{placebo}} &= 1 - p_1 \quad E_{\max} = p_1 - p_2 \\
\log\left(ED_{50}\right) &\sim N\left(0, 10^6\right) \quad \log\left(\text{base}\right) \sim N\left(0, 10^6\right) \\
\omega &\sim U\left(0, 10^4\right)
\end{aligned}
$$

WinBUGS implementation

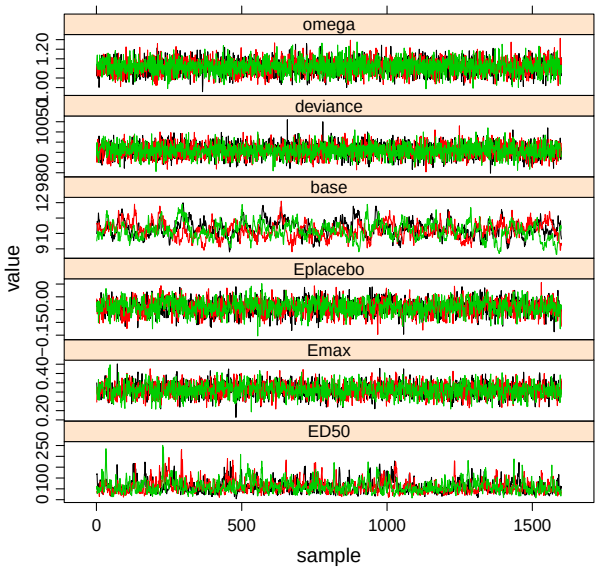
```
model{

  for(i in 1:nPatients){

    eta[i] ~ dnorm(0,tau)
    etaPred[i] ~ dnorm(0,tau)

  }

  for(i in 1:nObs){

    nSeizure[i] ~ dpois(lambda[i])
    nSeizureCond[i] ~ dpois(lambda[i])
    lambda[i] <- base * exp(eta[patient[i]]) *
      (1 - (1 - equals(time[i],0)) * (Edrug[i] + Eplacebo))
    Edrug[i] <- Emax*dose[i]/(ED50+dose[i])

    nSeizurePred[i] ~ dpois(lambdaPred[i])
    lambdaPred[i] <- base * exp(etaPred[patient[i]]) *
      (1 - (1 - equals(time[i],0)) * (Edrug[i] + Eplacebo))

  }
```

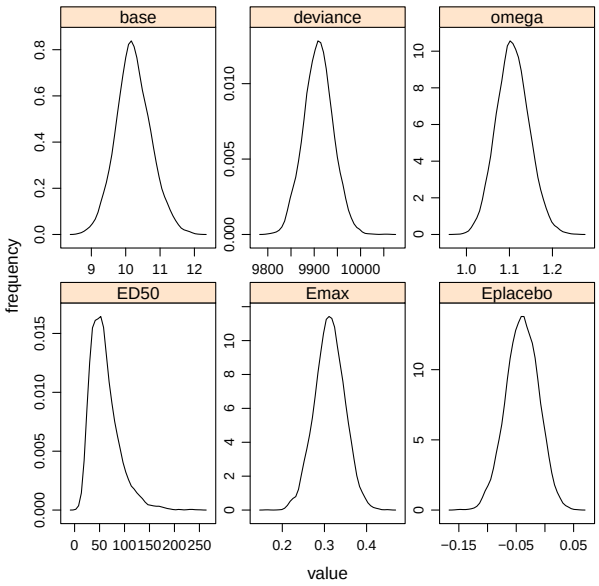
WinBUGS implementation

```
## require that Emax + Eplacebo < 1 and Eplacebo < 1

 logP1 ~ dnorm(0,1.0E-6)
 logP2 ~ dnorm(0,1.0E-6)
 log(p1) <- logP1
 log(p2) <- logP2

 Eplacebo <- 1 - p1
 Emax <- p1 - p2

 logED50 ~ dnorm(0,1.0E-6)
 log(ED50) <- logED50
 logBase ~ dnorm(0,1.0E-6)
 log(base) <- logBase
 omega ~ dunif(0,1.0E4)
 tau <- 1/(omega*omega)

}
```


Results

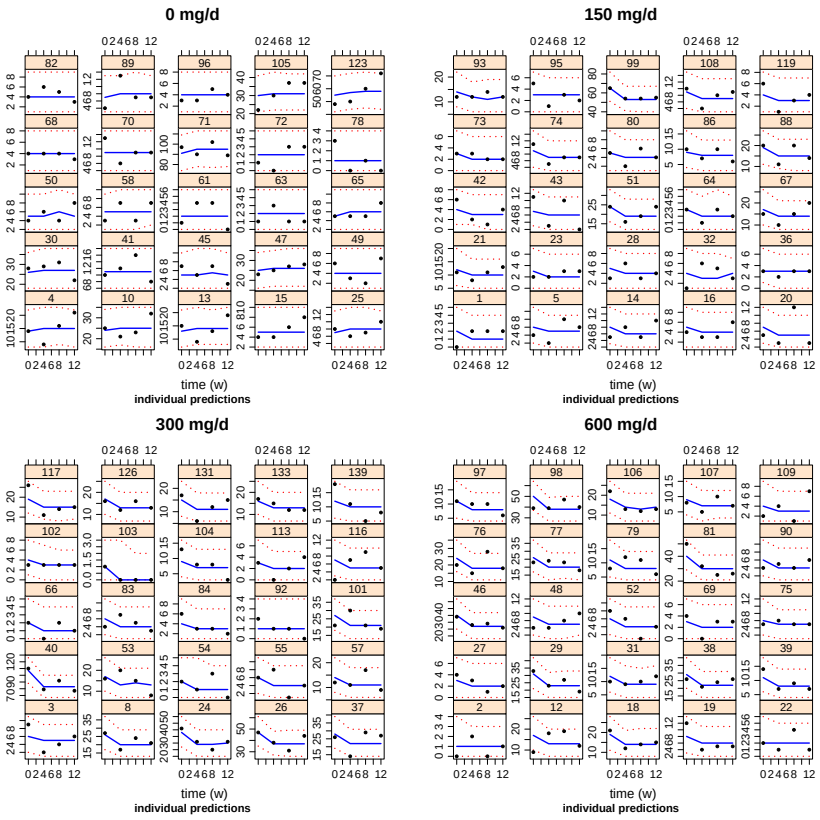
- 50000 × 3 chains
- burn-in = 10000 / chain, thin by 25





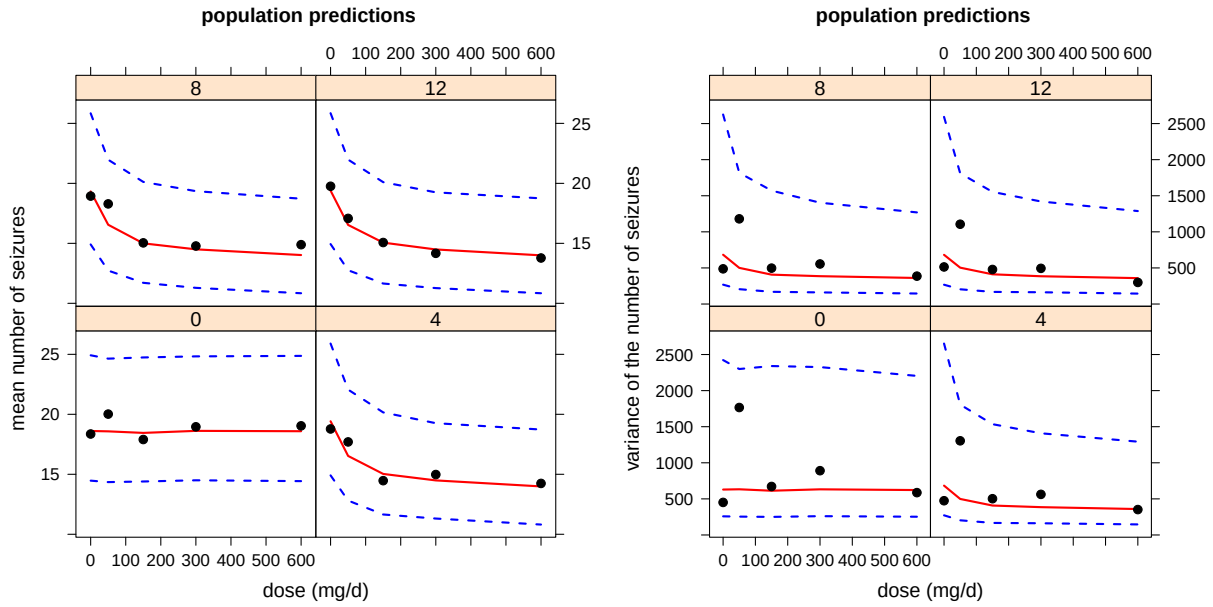
Model parameter estimates

parameter	mean	sd	2.5%ile	25%ile	median	75%ile	97.5%ile	effective N
deviance	9910	31.7	9850	9890	9910	9930	9970	5020
base	10.2	0.51	9.26	9.89	10.2	10.6	11.3	251
E_{placebo}	−0.0397	0.0281	−0.0965	−0.0589	−0.0394	−0.0202	0.0132	1640
$E_{\max}$	0.312	0.0352	0.244	0.289	0.312	0.336	0.381	2470
ED_{50}	60.2	29.2	21.6	39.2	54.8	74	134	734
ω	1.11	0.0379	1.04	1.08	1.11	1.13	1.18	4800



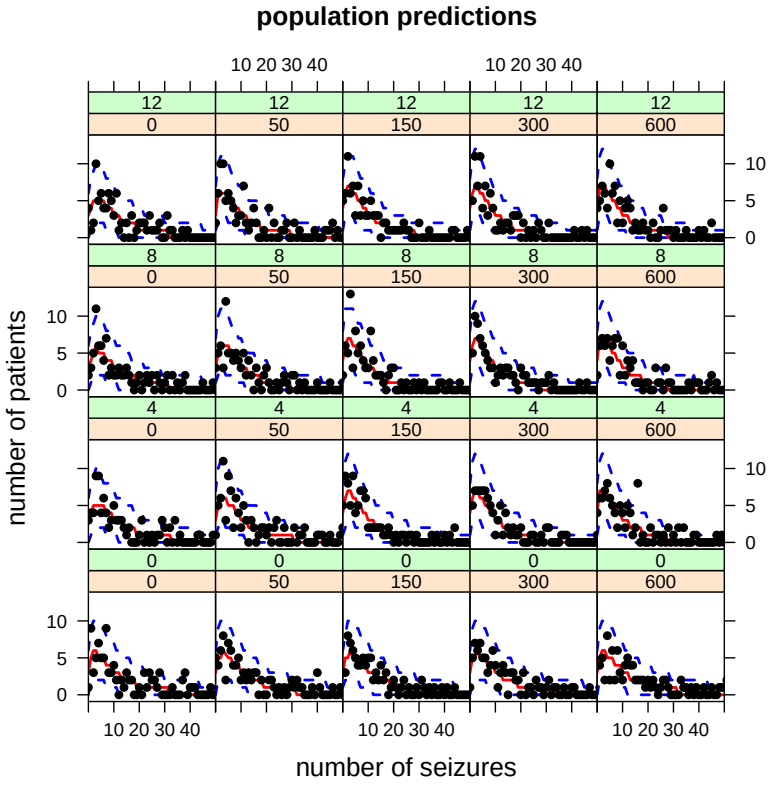
Representative individual fits, i.e., posterior predictions of new observations in the same patients (median & 90% prediction intervals)

Posterior predictions of sample means and variances



Posterior medians & 90% prediction intervals

Model predicted distribution of seizure counts



Posterior medians & 90% prediction intervals

Time-varying hazard model, a.k.a., non-homogeneous Poisson process

- The Poisson model may be extended to cases where the hazard varies systematically with time. This is known as a non-homogeneous Poisson process.
- If the hazard (expected event rate) is a function of time $h(t)$ then the number of events in any time interval (t_{i-1}, t_i) is a random variable distributed according to:

$$\Pr\left(N\left(t_i\right) - N\left(t_{i-1}\right) = n\right) = \frac{1}{n!}\left(\int_{t_{i-1}}^{t_i} h\left(t\right) dt\right)^n e^{-\int_{t_{i-1}}^{t_i} h(t)dt}$$

or equivalently

$$N\left(t_i\right) - N\left(t_{i-1}\right) \sim \text{Poisson}\left(\int_{t_{i-1}}^{t_i} h\left(t\right) dt\right)$$

Variations on the Poisson model to deal with over-dispersion or “zero inflation”

- For a Poisson random variable the mean and variance are equal.
- Suppose your data consist of event counts over the same length of time for a group of patients.
- If the expected event count is the same for all of the patients then the data are distributed according to a common Poisson distribution.
- The sample mean and variance will be approximately equal.
- But what if each individual’s event count is a sample from a Poisson distribution, but the expected event count varies among individuals?
  - Then the sample variance will usually be greater than the sample mean and the data are not consistent with a common Poisson distribution.
  - This is called over-dispersion
  - The Poisson model can be extended to account for such over-dispersion by using continuous mixture models in which the hazard is randomly distributed.

Variations on the Poisson model to deal with over-dispersion or “zero inflation”

What if the patients are drawn from 2 different populations, one of which has an expected event count of 0.

- That too will usually result in a sample variance that exceeds the sample mean, but for a very different reason.
- There will be an excess of 0’s in the data compared to Poisson distributed data.
- The Poisson model can be extended to this case by using a discrete mixture of a Poisson model plus a simple categorical model for the probability of no events—a so-called ZIP or Zero Inflated Poisson model.

Let’s use Hands-on Problem 4 to illustrate these models...

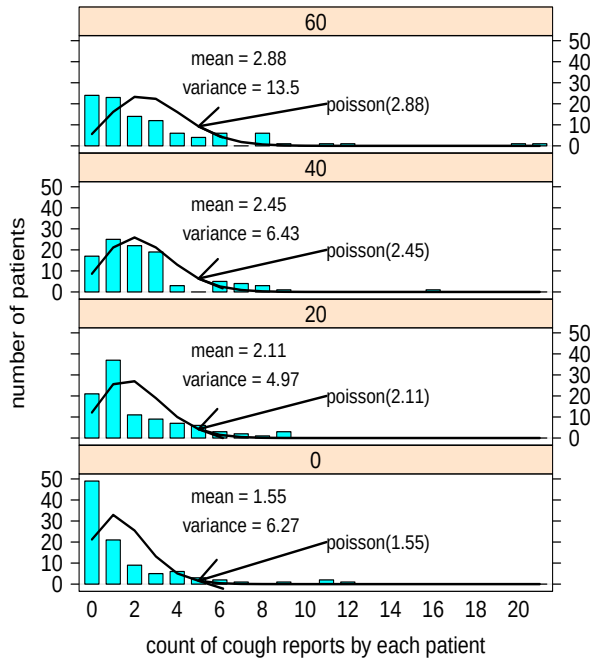
Hands-on problem 4: Count data

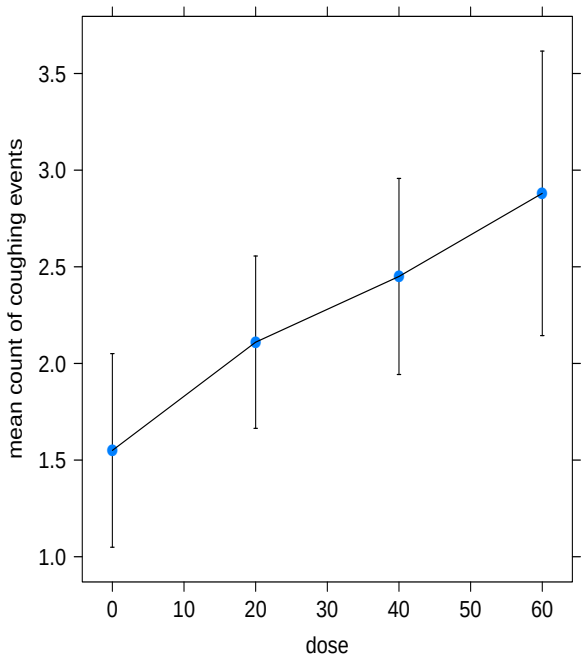
- Patients in the Phase 2 dose-finding study recorded selected efficacy and safety-related outcomes in a daily diary.
- This included an indication of whether they experienced excessive coughing.
- We will model the total number of coughing events in each patient as a function of dose.
- Focus on exploring different count models. This is a good setting for hands-on learning about model evaluation and selection.

Hands-on problem 4: Count data

- Phase 2 dose-finding trial in CF patients (same as hands-on problem 1)
  - Parallel design
  - 100 patients per dose arm
  - Multiple doses of ME-2
    - Placebo, 20, 40 and 60 mg qd administered by inhalation for 24 weeks
  - AE measurement: Total number of coughing events.
- Hands-on exercise:
  - Construct a model for the count of coughing events as a function of dose.
- Data file: handsOn4/ME2CoughCountData.csv

EDA: Distribution of cough event counts by dose





Proposed models

Simple Poisson model:

$$\begin{aligned} n_{\text{cough},i} &\sim \text{Poisson}\left(\lambda_i t_i\right) \\ \lambda_i &= a + bD_i \end{aligned}$$

where t_i is the duration over which the coughing events were observed.

Weakly informative priors:

$$\log\left(a\right) \sim N\left(0, 10^6\right) \quad \log\left(b\right) \sim N\left(0, 10^6\right)$$

Proposed models

Negative binomial model (Poisson-gamma mixture) assuming the standard deviation of cough frequency is proportional to the mean:

$$\begin{aligned} n_{\text{cough},i} &\sim \text{Poisson}\left(\lambda_i t_i\right) \\ \lambda_i &= \text{gamma}\left(\alpha, \beta_i\right) \\ \beta_i &= \frac{\alpha}{\widehat{\lambda}_i} \\ \widehat{\lambda}_i &= a + bD_i \end{aligned}$$

Weakly informative priors:

$$\log\left(a\right) \sim N\left(0, 10^6\right) \quad \log\left(b\right) \sim N\left(0, 10^6\right) \quad \log\left(\alpha\right) \sim N\left(0, 10^6\right)$$

The model is parameterized such that $E\left(n_{\text{cough},i}\right) = \widehat{\lambda}_i t_i$ and $\text{Var}\left(n_{\text{cough},i}\right) = \frac{\left(\widehat{\lambda}_i t_i\right)^2}{\alpha}$, i.e., the standard deviation is proportional to the mean.

Proposed models

Poisson-lognormal mixture model assuming the standard deviation of cough frequency is proportional to the mean:

$$\begin{aligned} n_{\text{cough},i} &\sim \text{Poisson}\left(\lambda_i t_i\right) \\ \log\left(\lambda_i\right) &= N\left(\log\left(\widehat{\lambda}_i\right), \sigma^2\right) \\ \widehat{\lambda}_i &= a + bD_i \end{aligned}$$

Weakly informative priors:

$$\log\left(a\right) \sim N\left(0, 10^6\right) \quad \log\left(b\right) \sim N\left(0, 10^6\right) \quad \sigma \sim U\left(0, 10^4\right)$$

Proposed models

Zero-inflated Poisson (ZIP) model:

$$n_{\text{cough},i} \sim \text{ZIP}\left(\lambda_i t_i, \pi\right)$$

where

$$\begin{aligned} p_{\text{ZIP}}\left(x|\lambda_i t_i, \pi\right) &= \left(1-\pi\right) p_{\text{Poisson}}\left(x|0\right) + \pi p_{\text{Poisson}}\left(x|\lambda_i t_i\right) \\ &= \begin{cases} 1 - \pi + \pi p_{\text{Poisson}}\left(0|\lambda_i t_i\right), & x = 0 \\ \pi p_{\text{Poisson}}\left(x|\lambda_i t_i\right), & x > 0 \end{cases} \\ \lambda_i &= a + bD_i \end{aligned}$$

Weakly informative priors:

$$\log\left(a\right) \sim N\left(0, 10^6\right) \quad \log\left(b\right) \sim N\left(0, 10^6\right) \quad \pi \sim U\left(0, 1\right)$$

Proposed models

Alternative representation of the zero-inflated Poisson (ZIP) model:

$$n_{\text{cough},i} \sim \text{Poisson}\left(\lambda_{\text{ZIP},i} t_i\right)$$

where

$$\begin{aligned} \lambda_{\text{ZIP},i} &= I_{\text{Poisson},i}\lambda_i + \left(1 - I_{\text{Poisson},i}\right) 0 \\ I_{\text{Poisson},i} &\sim \text{Bernoulli}\left(\pi\right) \\ \lambda_i &= a + bD_i \end{aligned}$$

Weakly informative priors:

$$\log\left(a\right) \sim N\left(0, 10^6\right) \quad \log\left(b\right) \sim N\left(0, 10^6\right) \quad \pi \sim U\left(0, 1\right)$$

Hands-on problem 4: Model comparison

Model	E (deviance)	pD	DIC
Poisson	1922	1.95	1924
Negative binomial (Poisson-gamma)	1201	198	1399
Poisson-lognormal	1219	215	1434
ZIP	1490	411	1900

Modeling time-to-event data for a single event per individual

Examples of time-to-event data in clinical pharmacology

Single events

- Morbidity and mortality, e.g., death, heart attack, stroke, ...
- Dropout from study
- Thrombosis
- Bleeding
- First occurrence of what may be multiple events, e.g., AE's

Multiple events

- AE's
- Seizures
- Vomiting
- Migraine
- Dosing events (when modeling adherence)

What makes time-to-event data "odd"?

- Time-to-event measurements are continuous data, so why don't we analyze them just like any other continuous PD measurements?
- Two features distinguish time-to-event data from most common PD measurements:
  - Most time-to event data sets include censored data, particularly right censored data.
    - Right censoring refers to the case where no event occurs in an individual during the period of observation.
    - You don't know the time to the event, but you do know that the time to the event must exceed the duration of the observation period.
  - Time-to-event data are not observed at some pre-specified observation time and they reflect the risk of an event over the entire observation period up to the time the event occurs—not just the risk at the time of the event.

Principles and methods of survival analysis for modeling censored data

Concept and formal definition of hazard

- Let T be a random variable representing the time to some event.
- t represents elapsed time since some specified starting event, usually the start of a study.
- The basic idea is to model the probability distribution of T as a function of various covariates, e.g., time, dose, pharmacologic response, patient characteristics, etc.
- Such models are often most naturally conceptualized in terms of the "hazard" rate.
- Hazard may be interpreted as the instantaneous probability density (probability per unit time) of an event occurring given that it has not yet occurred.

The hazard function:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Relationship of hazard to probability distribution of event times

Survival function

$$S(t) = \Pr(T > t) = e^{-\int_0^t h(u)du}$$

Cumulative distribution function (c.d.f)

$$F(t) = \Pr(T \leq t) = 1 - S(t) = 1 - e^{-\int_0^t h(u)du}$$

Probability density

$$f(t) = F'(t) = h(t)\,S(t)$$

Likelihood function for time-to-event data

Likelihood for times of observed events, i.e., it happened at time t:

$$p(t|\theta) = f(t|\theta) = h(t|\theta)\,S(t|\theta)$$

Likelihoods for censored time-to-event data

- Right censored data, i.e., it hasn't happened up to time t:

$$p(t|\theta) = \Pr(T > t|\theta) = 1 - F(t|\theta) = S(t|\theta)$$

- Interval censored data, i.e., it happened sometime between time t_1 and time t_2:

$$p(t|\theta) = \Pr(t_1 < T \leq t_2|\theta) = F(t_2|\theta) - F(t_1|\theta) = S(t_1|\theta) - S(t_2|\theta)$$

Dealing with censored data in WinBUGS

- Right, left and interval censored data are easily handled in WinBUGS by modifying the likelihood distribution with I(lower,upper). Lower or upper may be blank meaning no limit. For example:
  - Suppose y[i] is normally distributed with some mean mu[i] and precision tau and
  - A point value of y[i] is not observed but
  - You know it lies between 5 and 10
  - Then we say it is interval censored
  - The likelihood may described in WinBUGS by
    - y[i] $\sim$ dnorm(mu[i],tau)I(5,10)
    - The corresponding y[i] should be NA in the data set.

Models with constant hazard

If the hazard function is constant with respect to time, i.e., $h(t) = h$, the following simplified relations result:

- Survival function

$$S(t) = \Pr(T > t) = e^{-ht}$$

- Cumulative distribution function (c.d.f)

$$F(t) = \Pr(T \leq t) = 1 - S(t) = 1 - e^{-ht}$$

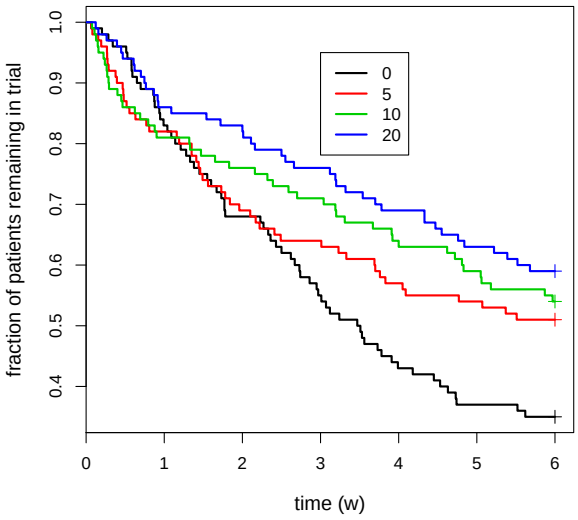
- Probability density

$$f(t) = F'(t) = hS(t) = he^{-ht}$$

In other words the event time T is distributed according to an exponential distribution.

Example: Modeling dropouts: Constant hazard model

- Consider the problem of modeling dropout behavior for a hypothetical antipsychotic drug candidate.
- Phase 2 dose-finding trial
- 4 treatment arms: placebo and 5, 10 and 20 mg/d
- 100 patients per arm
- Efficacy assessed based on weekly PANSS scores



Proposed initial model

Let's start with a simple constant hazard base model that depends only on the daily dose:

$$\begin{aligned} t_{\text{drop},i} &\sim \text{Exponential}\left(h_{\text{drop},i}\right) \\ h_{\text{drop},i} &= h_0 e^{\beta D} \end{aligned}$$

Uninformative priors:

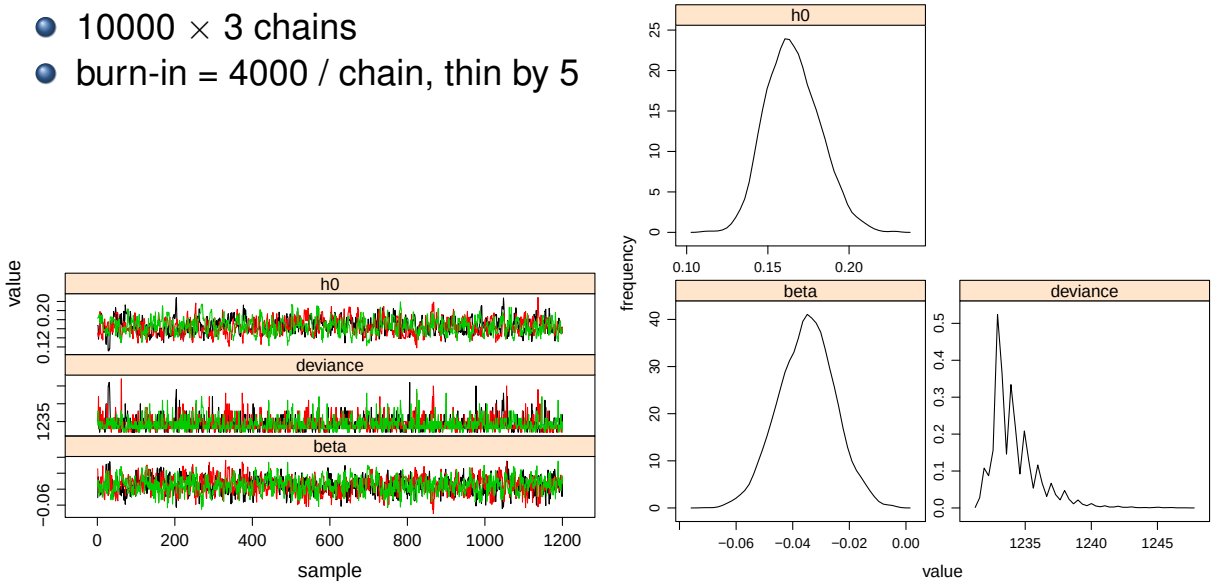
$$\log\left(h_o\right) \sim N\left(0, 10^6\right) \qquad \beta \sim N\left(0, 10^6\right)$$

WinBUGS implementation

```
model{

  for(i in 1:nPatients){

    tDrop[i] ~ dexp(h[i])I(tCensor[i],)
    h[i] <- h0 * exp(beta * dose[i])

    tDropPred[i] ~ dexp(h[i])

  }

  logH0 ~ dnorm(0, 1.0E-6)
  beta ~ dnorm(0, 1.0E-6)
  log(h0) <- logH0

}
```

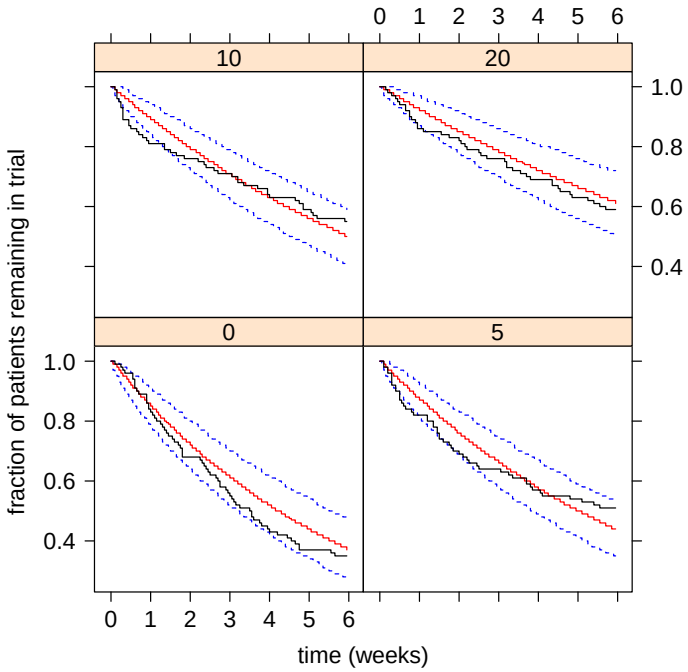

Results

- 10000 × 3 chains
- burn-in = 4000 / chain, thin by 5



parameter	mean	sd	2.5%ile	25%ile	median	75%ile	97.5%ile	effective N
deviance	1230	1.88	1230	1230	1230	1240	1240	1830
h_0	0.166	0.0165	0.136	0.154	0.164	0.176	0.2	776
β	−0.0351	0.00987	−0.0545	−0.0418	−0.0349	−0.0284	−0.0156	987

Posterior median and 90% prediction interval for the fraction of patients remaining in the trial compared to the empirical survival curve



Hands-on problem 5: Time-to-event data

- ME-3, a more potent analog of ME-2, is also a mucolytic that is being developed as a treatment for cystic fibrosis.
- Phase I trials are in progress, so the current focus is on the design of Phase 2 PoC and dose-finding trials.
- The development team's preferred primary endpoint for efficacy assessment is time to the first pulmonary exacerbation event, but trials using conventional hypothesis tests require large sample sizes and/or durations in order to achieve adequate statistical power.
- A model relating FEV1 to exacerbation hazard was developed by meta-analysis of results from past ME-2 trials plus summary data for other mucolytics reported in public sources.
  - It is believed that the model is qualitatively and quantitatively applicable to the new drug candidate because the drugs work by the same mechanism and the patient population is essentially the same
  - E.g., standard of care has not changed since the previous trials.

Hands-on problem 5: Time-to-event data

- The team wants to make PoC and dose selection decisions based on pulmonary exacerbations but it also hopes to accelerate the development program.
- It may be possible to reduce sample sizes or trial duration for a Phase 2 PoC and dose-finding trial by analyzing both exacerbation and FEV1 data using a joint model that incorporates the aforementioned model relating FEV1 to exacerbation hazard—including the informative prior distribution of its parameters.
- This would permit inferences regarding exacerbations conditioned on prior knowledge and the observed FEV1 and pulmonary exacerbation data.
- Based on this idea the team conducted a study with fewer patients and half the treatment duration (12 weeks) than a typical study for a CF mucolytic.
- We analyze the results in this hands-on example.

Hands-on problem 5: Time-to-event data

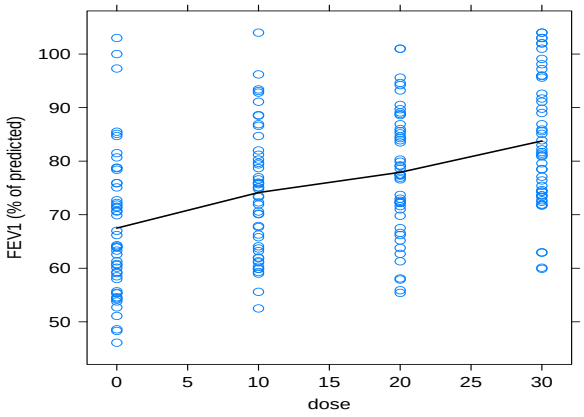
Phase 2 dose-finding trial in CF patients

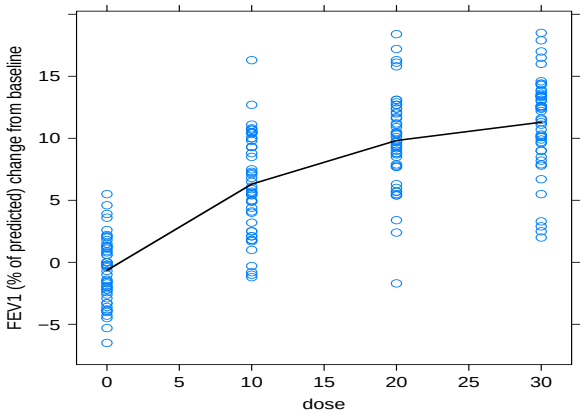
- Parallel design
- 50 patients per dose arm
- Multiple doses of ME-3
  - Placebo, 10, 20 and 30 mg qd administered by inhalation for 12 weeks
- Primary efficacy measurement: Time to first pulmonary exacerbation event within 12 weeks
- Secondary/supportive measurement: FEV1 at 12 weeks

Hands-on exercise:

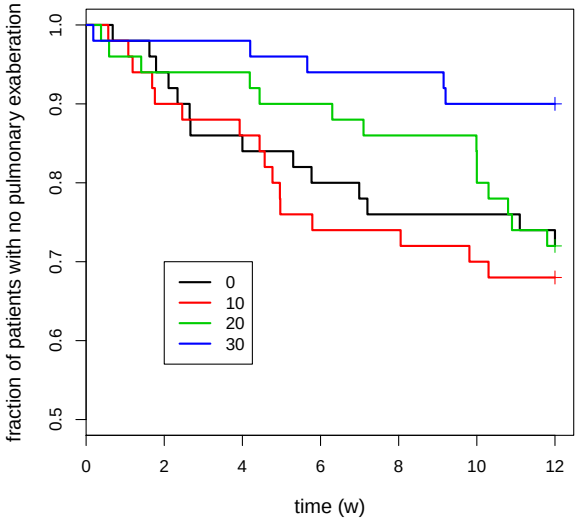
- Construct a joint model for time to first pulmonary exacerbation and FEV1 as a function of dose conditioned on a previously developed model for the relationship between FEV1 and pulmonary exacerbation hazard including an informative prior distribution for the parameters of that model.

Data file: handsOn5/ME3ExacerbationFEV1Data.csv





EDA: Time to first pulmonary exacerbation and FEV1 as a function of dose



Proposed model

- Constant hazard model for occurrence of pulmonary exacerbation in the i^{th} patient where the hazard is a function of model-predicted FEV1.
- Emax model for FEV1 with log normally distributed residual variation.

$$\begin{aligned} t_{\text{exac},i} &\sim \text{Exponential}\left(h_i\right) \\ \log\left(FEV1_i\right) &\sim N\left(\log\left(\widehat{FEV1}_i\right), \sigma^2_{FEV1}\right) \\ h_i &= h_0 e^{-\beta\left(\widehat{FEV1}_i - 70\right)} \\ \widehat{FEV1}_i &= FEV1_{0i} + \frac{E_{\max} D_i}{ED_{50} + D_i} \end{aligned}$$

where $FEV1_{0i}$ is the observed baseline FEV1.

- Informative prior distribution for h_0 and β:

$$\log\left(h_0\right) \sim N\left(\log\left(0.029\right), 0.25^2\right) \quad \beta \sim N\left(0.031, 0.015^2\right)$$

- Weakly informative priors for the remaining parameters:

$$\log\left(E_{\max}\right) \sim N\left(0, 10^6\right) \quad \log\left(ED_{50}\right) \sim N\left(0, 10^6\right) \quad \sigma_{FEV1} \sim U\left(0, 10^4\right)$$

PII-108

"TRUNCATED SIGMOID E_{max} MODELS": A REPARAMETERIZATION OF THE SIGMOID E_{max} MODEL FOR USE WITH TRUNCATED PK/PD DATA. WJ Bachman PhD and WR Gillespie PhD, GloboMax LLC, Hanover, MD.

The parameters of the sigmoid E_{max} model are poorly estimated when the range of PK/PD data available is limited to $<0.95E_{max}$ [Dutta et al. J Pharm Sci 85:232 (1996)]. The following reparameterized form of the sigmoid E_{max} model has improved parameter estimation properties:

$$E = E_0 + \frac{(\beta^\gamma + 1)(E^* - E_0)C^\gamma}{C^{*\gamma} + \beta^\gamma C^\gamma}$$

where E is the effect measure and C is a measure of drug exposure (e.g., concentration or dose). The parameter E^* is the estimated effect resulting from C^*, γ is the usual "sigmoidicity" parameter, and E_0 is the baseline effect. β is a measure of the degree to which the function deviates from linearity in C^γ. One approach to applying this parameterization is to fix C^* (or E^*) at a value and estimate the remaining parameters E_0, E^* (or C^*), β, and γ by nonlinear regression. The properties of this approach are evaluated by application to simulated PK/PD data that is truncated at various fractions of E_{max}. When C^* (or E^*) is chosen within the range of the observed data, then the parameters E^* (or C^*) and β are more precisely and accurately estimated than EC_{50} and E_{max} of the standard parameterization.

- Model parameterization implemented in handsOn5TruncEmax.txt
- CP&T 63:199 (1998)

Models with time-varying hazard

Time-to-event models based on time-varying hazard functions may be appropriate in some cases, e.g.,

- The risk of an event at a particular time is a function of the magnitude of drug exposure at that time and drug exposure is varying over time.
- Some form of tolerance occurs that results in a reduced incidence of events over time.
- The risk of an event is causally related to some observable response that varies with time, e.g., risk of dropout may be increased by perceived lack of efficacy.

Models with time-varying hazard

Implementation requires integration of the hazard function within the likelihood function:

- Likelihood for times of observed events

$$p(t|\theta) = f(t|\theta) = h(t|\theta)\, S(t|\theta) = h(t|\theta)\, e^{-\int_0^t h(u|\theta)du}$$

- Likelihoods for censored time-to-event data
  - Right censored data

$$p(t|\theta) = \Pr(T > t|\theta) = 1 - F(t|\theta) = S(t|\theta) = e^{-\int_0^t h(u|\theta)du}$$

  - Interval censored data

$$\begin{aligned} p(t|\theta) &= \Pr(t_1 < T \leq t_2|\theta) = F(t_2|\theta) - F(t_1|\theta) \\ &= S(t_1|\theta) - S(t_2|\theta) = e^{-\int_0^{t_1} h(u|\theta)du} - e^{-\int_0^{t_2} h(u|\theta)du} \end{aligned}$$

Implementing models with time-varying hazard

NONMEM implementation:

- Use PREDPP models for numerical integration, e.g., ADVAN6, ADVAN8 or ADVAN13.

WinBUGS implementation:

- Use one of the ODE solvers in BUGSModelLibrary (http://bugsmodellibrary.googlecode.com) to integrate the hazard function.
- And use the zeros or ones trick to specify the non-standard sampling distribution resulting from the time-varying hazard.

Implementing models with time-varying hazard

If the hazard function can be described (approximated) as piecewise constant then:

- The duration of each constant hazard interval prior to an event and during which no event occurs is treated as a right censored time-to-event from an exponential distribution.
- For the constant hazard interval during which an event occurs, the time from the start of the interval to the time of the event is modeled as an exponential random variable.

Time-varying hazard models pose major identifiability problems when modeling time to a single event.

- It is often not possible to distinguish between a constant hazard model and a time-varying hazard model with the same time-averaged hazard.

Example: Piecewise constant hazard model for dropouts

- Let's revisit the dropout example.
- It is likely that dropout behavior is influenced by both efficacy (as measured by PANSS) and adverse effects (e.g., CNS and extrapyramidal symptoms (EPS)).
- To keep things simple let's just model the effect of lack of efficacy on dropout. In particular construct a model where the hazard depends on the PANSS score observed at the most recent weekly visit.
- Model the hazard as constant over the period since the most recent PANSS score measurement.

WinBUGS implementation

The BUGS model is nearly identical to the previous one except that the covariate is now PANSS change from baseline instead of dose.

```
model{

  for(i in 1:nObs){

    tDrop[i] ~ dexp(h[i])I(tCensor[i],)
    h[i] <- h0 * exp(beta * (PANSS[i]-PANSS0[i]))

    tDropPred[i] ~ dexp(h[i])

  }

  logH0 ~ dnorm(0, 1.0E-6)
  beta ~ dnorm(0, 1.0E-6)
  log(h0) <- logH0

}
```


WinBUGS implementation

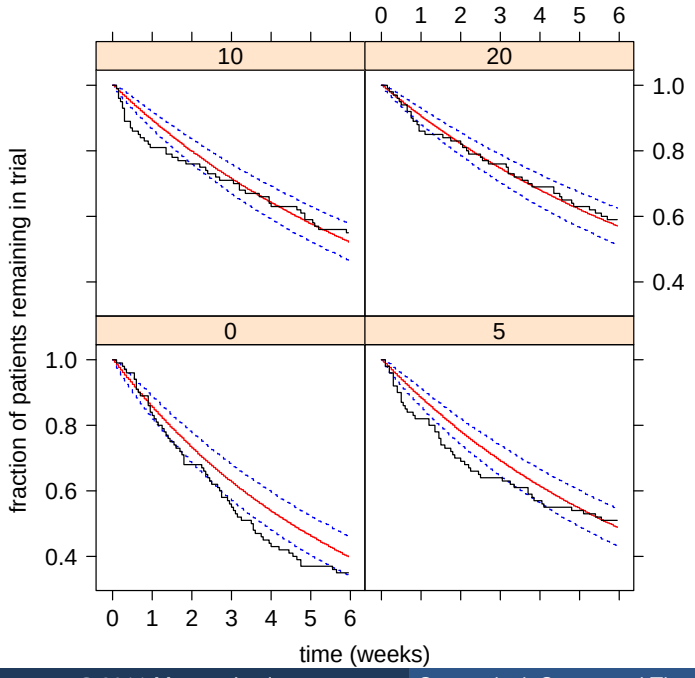
The bigger difference is in the data set that now contains a tDrop record indicating the dropout (or censored if 6) time.

patient	dose	PANSS0	time	PANSS	tDrop
1	0	90	0	91	1.77
1	0	90	1	87	1.77
1	0	90	2	NA	1.77
1	0	90	3	NA	1.77
1	0	90	4	NA	1.77
1	0	90	5	NA	1.77
1	0	90	6	NA	1.77
2	5	79	0	78	6
2	5	79	1	71	6
2	5	79	2	74	6
2	5	79	3	66	6
2	5	79	4	58	6
2	5	79	5	65	6
2	5	79	6	69	6

The data is read in and reformatted by the following R code

```
## get data file
dataTTE <- read.csv("PANSSData.csv",as.is=T)

dataTTE$tEnd <- as.vector(
  sapply(unique(dataTTE$patient),
         function(patient, data)
         c(data$time[data$patient==patient][-1], NA),
         data = dataTTE))

dataTTE <- dataTTE[!is.na(dataTTE$PANSS) & !is.na(dataTTE$tEnd), ]
dataTTE$censored <- dataTTE$tDrop >= dataTTE$tEnd
dataTTE$tEnd <- ifelse(dataTTE$tDrop < dataTTE$tEnd, dataTTE$tDrop,
  dataTTE$tEnd)

## create WinBUGS data set
bugsdata <- list(
  nObs = nrow(dataTTE),
  PANSS = dataTTE$PANSS,
  PANSS0 = dataTTE$PANSS0,
  tDrop = ifelse(dataTTE$censored, NA, dataTTE$tEnd - dataTTE$time),
  tCensor = ifelse(!dataTTE$censored, 0, dataTTE$tEnd - dataTTE$time)
  )
```


Results: Posterior median and 90% prediction interval for the fraction of patients remaining in the trial compared to the empirical survival curve



The fit is only slightly better than the constant hazard dose-response model based on expected deviance (1210 vs. 1230).

Modeling repeated time-to-event data

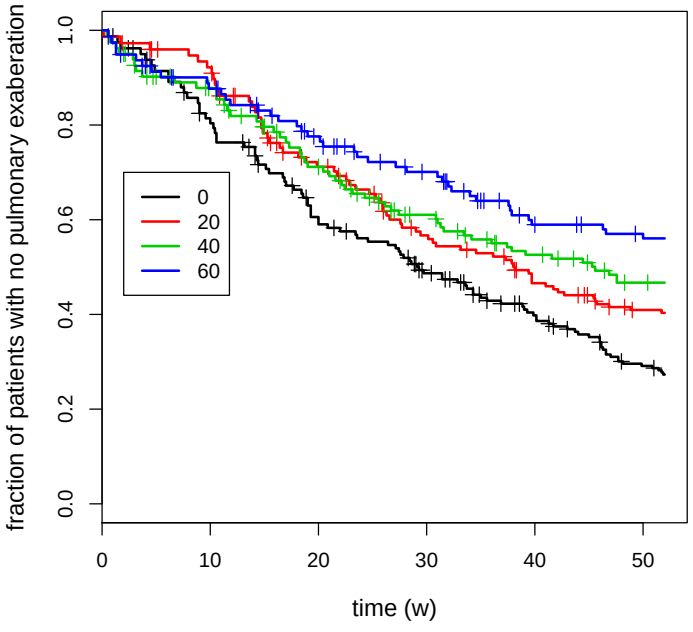
Approaches to modeling multiple events of the same kind

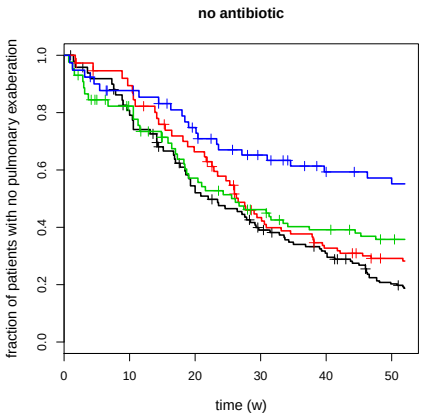
- Model the time to the first event
- Model the number of events
- Model the inter-event time intervals
  - This is a relatively straight forward extension of the approach used for single events.
  - Instead of modeling only the time of an event relative to the start of the study, you model the time from the previous event (or the start of the study in the case of the first event).
  - This approach is potentially more informative than modeling count data when the hazard varies with time.
  - If sufficient individuals experience multiple events then modeling of inter-individual variation is possible.
- Let's use hands-on problem 6 to illustrate the approach

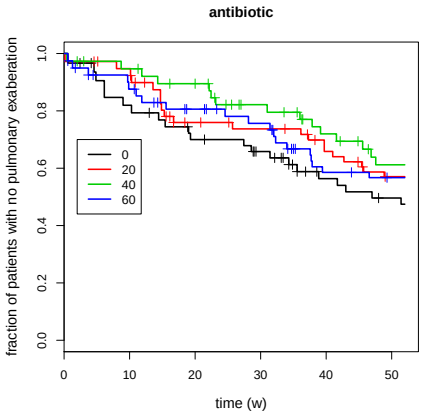
Hands-on Problem 6: Repeated time-to-event data

- Suppose the trial we modeled in Hands-on Problem 1 continued for 1 year instead of 24 weeks (an artifice to assure we have enough multiple events to make it interesting).
- Phase 2 dose-finding trial in CF patients
  - Parallel design
  - 100 patients per dose arm
  - Multiple doses of ME-2
    - Placebo, 20, 40 and 60 mg qd administered by inhalation for 52 weeks
  - Efficacy measurement: Times of pulmonary exacerbation events.
  - Covariates of possible relevance: age, baseline FEV1, concomitant medications (rhDNase or chronic antibiotic such as azithromycin or inhaled tobramycin)
- Hands-on exercise:
  - Construct a model for the possibly multiple times to pulmonary exacerbation events as a function of dose and possibly patient-specific covariates.
- Data file: handsOn6/ME2ExacerbationTimeData.csv

EDA: Time between pulmonary exacerbation events as a function of dose and antibiotic use







Proposed model

- Constant hazard model for i^{th} occurrence of a pulmonary exacerbation in the j^{th} patient where the hazard is a function of dose, baseline FEV1 and chronic antibiotic use:

$$\begin{aligned} t_{\text{exac},ij} &\sim \text{Exponential}\left(h_j\right) \\ h_j &= h_{0j} E_{\text{drug},j} \\ \log\left(h_{0j}\right) &\sim N\left(\log\left(\widehat{h}_{0j}\right), \omega^2\right) \\ \log\left(\widehat{h}_{0j}\right) &= \beta_0 + \beta_{FEV1}\left(FEV1_j - 70\right) + \beta_{\text{antibiotic}} I_{\text{antibiotic},j} \\ E_{\text{drug},j} &= 1 - \frac{E_{\max} D_j}{ED_{50} + D_j} \end{aligned}$$

  where $FEV1_j$ is the observed baseline FEV1.

- Weakly informative priors:

$$\begin{aligned} \beta_0 &\sim N\left(0, 10^6\right) \quad \beta_{FEV1} \sim N\left(0, 10^6\right) \quad \beta_{\text{antibiotic}} \sim N\left(0, 10^6\right) \\ E_{\max} &\sim U\left(0, 1\right) \quad ED_{50} \sim U\left(0, 1000\right) \quad \omega \sim U\left(0, 10^4\right) \end{aligned}$$

Possible additional topics

We will probably have time to cover 2 additional topics. Here are some possibilities:

- Modeling non-ordinal categorical data.
- Modeling ordinal data when proportional odds is not appropriate.
- Hidden Markov models—an approach for dealing with some types of autocorrelation.
- Simultaneous modeling of time-to-event and event magnitude.
- Models for time-to-event or count data with a hazard that varies continuously with time.
- Other?

Please indicate your priorities on the Student Q&A Forum of the course website.

Hidden Markov models

- Hidden Markov models are models in which the observable data depends on some underlying Markov process.
- The modeled system has multiple possible states, but can be in only one such state at a time.
- In a discrete time Markov process the time scale is a sequence of discrete times, usually equally spaced.
  - The system can transition randomly from one state to another at each time.
  - The probability of each transition depends only on the current state, not on any prior state. Thus the sequence of states is a Markov chain.
  - The transition probabilities form a transition matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ P_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix}$$

  where p_{ij} is the probability of transitioning from state i to state j.

Potential applications

- Describing some types of autocorrelation in discrete data models
  - Binary or binomial data in which the probability of an event changes at random from one constant value to another.
  - Count data where the event hazard changes at random from one constant value to another.

Example: Hidden Markov model for seizure count data

The example we will explore is based on the following work:

- M Delattre, R Savic, R Miller, M O Karlsson, M Lavielle. Estimation of mixed hidden Markov models with SAEM. Application to daily seizures data. ACCP Annual Meeting 2009 (`http://accp1.org/pdf/2009/Lavielle.pdf`)
- M Delattre, R Savic, R Miller, M O Karlsson, M Lavielle. Estimation of mixed hidden Markov models with SAEM. Application to daily seizures data. PAGE 2010 (`http://www.page-meeting.org/pdf_assets/6696-PAGE2010.pdf`)

Worked example using WinBUGS

Simulated seizure counts from a hypothetical clinical trial

- 5 treatment arms
  - 0, 600, 900, 1200 and 1800 mg/d
- 12 week screening phase followed by 12 week treatment phase
- 25 patients per arm
- Data: Daily seizure counts

Proposed model

Poisson model for the number of seizures on the i^{th} observation day in the j^{th} patient with average seizure counts that vary between 2 values according to a Markov process:

$$
\begin{aligned}
n_{\text{seizure},ij} &\sim \text{Poisson}\left(\lambda_{ij}\right) \\
\lambda_{ij} &= \begin{cases} \lambda_{1j}^{\text{screen}}, & \text{state}_{ij} = 1 \;\&\; t \leq 0 \\ \lambda_{2j}^{\text{screen}}, & \text{state}_{ij} = 2 \;\&\; t \leq 0 \\ \lambda_{1j}^{\text{treat}}, & \text{state}_{ij} = 1 \;\&\; t > 0 \\ \lambda_{2j}^{\text{treat}}, & \text{state}_{ij} = 2 \;\&\; t > 0 \end{cases} \\
\text{state}_{ij} &\sim \text{categorical}\left(p_{ij}\right) \\
p_{ij} &= \begin{cases} \left(p_{11,j}^{\text{screen}}, p_{12,j}^{\text{screen}}\right), & \text{state}_{(i-1)j} = 1 \;\&\; t \leq 0 \\ \left(p_{21,j}^{\text{screen}}, p_{22,j}^{\text{screen}}\right), & \text{state}_{(i-1)j} = 2 \;\&\; t \leq 0 \\ \left(p_{11,j}^{\text{treat}}, p_{12,j}^{\text{treat}}\right), & \text{state}_{(i-1)j} = 1 \;\&\; t > 0 \\ \left(p_{21,j}^{\text{treat}}, p_{22,j}^{\text{treat}}\right), & \text{state}_{(i-1)j} = 2 \;\&\; t > 0 \end{cases}
\end{aligned}
$$

Proposed model (cont.)

$$
\begin{aligned}
&\left(\text{logit}\left(p_{11,j}^{\text{screen}}\right), \text{logit}\left(p_{21,j}^{\text{screen}}\right), \log\left(\lambda_{1j}^{\text{screen}}\right), \log\left(\lambda_{2j}^{\text{screen}} - \lambda_{1j}^{\text{screen}}\right)\right) \sim \\
&\qquad N\left(\left(\text{logit}\left(\widehat{p}_{11}\right), \text{logit}\left(\widehat{p}_{21}\right), \log\left(\widehat{\lambda}_1\right), \log\left(\widehat{\Delta\lambda}\right)\right), \Omega\right) \\
p_{12,j}^{\text{screen}} &= 1 - p_{11,j}^{\text{screen}} \qquad p_{22,j}^{\text{screen}} = 1 - p_{21,j}^{\text{screen}} \\
\text{logit}\left(p_{11,j}^{\text{treat}}\right) &= \text{logit}\left(p_{11,j}^{\text{screen}}\right) + \delta_1 + \gamma_1 * D_j \\
\text{logit}\left(p_{21,j}^{\text{treat}}\right) &= \text{logit}\left(p_{21,j}^{\text{screen}}\right) + \delta_2 + \gamma_2 * D_j \\
p_{12,j}^{\text{treat}} &= 1 - p_{11,j}^{\text{treat}} \qquad p_{22,j}^{\text{treat}} = 1 - p_{21,j}^{\text{treat}} \\
\log\left(\lambda_{1j}^{\text{treat}}\right) &= \log\left(\lambda_{1j}^{\text{screen}}\right) + \delta_3 + \gamma_3 * D_j \\
\log\left(\lambda_{2j}^{\text{treat}} - \lambda_{1j}^{\text{treat}}\right) &= \log\left(\lambda_{2j}^{\text{screen}} - \lambda_{1j}^{\text{screen}}\right) + \delta_4 + \gamma_4 * D_j
\end{aligned}
$$

What about the initial state?

In the model described above the probability of each state depends on the immediately preceding state. How do we model the probability of each state at the first observation time? Here are 2 options:

1. Add new parameters corresponding to the initial probabilities.
2. Assume that the Markov process in this model has continued long enough prior to the study in conditions comparable to the screening phase that its stationary distribution may be used to estimate the probabilities of each state:

$$p_{1j} = \left(\frac{p_{21,j}^{\text{screen}}}{p_{12,j}^{\text{screen}} + p_{21,j}^{\text{screen}}}, \frac{p_{12,j}^{\text{screen}}}{p_{12,j}^{\text{screen}} + p_{21,j}^{\text{screen}}} \right)$$

Models with time-varying hazard (reprise)

Implementation requires integration of the hazard function within the likelihood function:

- Likelihood for times of observed events

$$p(t|\theta) = f(t|\theta) = h(t|\theta)\, S(t|\theta) = h(t|\theta)\, e^{-\int_0^t h(u|\theta)du}$$

- Likelihoods for censored time-to-event data
  - Right censored data

$$p(t|\theta) = \Pr(T > t|\theta) = 1 - F(t|\theta) = S(t|\theta) = e^{-\int_0^t h(u|\theta)du}$$

  - Interval censored data

$$\begin{aligned} p(t|\theta) &= \Pr(t_1 < T \leq t_2|\theta) = F(t_2|\theta) - F(t_1|\theta) \\ &= S(t_1|\theta) - S(t_2|\theta) = e^{-\int_0^{t_1} h(u|\theta)du} - e^{-\int_0^{t_2} h(u|\theta)du} \end{aligned}$$

Implementing models with time-varying hazard (reprise)

NONMEM implementation:

- Use PREDPP models for numerical integration, e.g., ADVAN6, ADVAN8 or ADVAN13.

WinBUGS implementation:

- Use one of the ODE solvers in BUGSModelLibrary (http://bugsmodellibrary.googlecode.com) to integrate the hazard function.
- And use the zeros or ones trick to specify the non-standard sampling distribution resulting from the time-varying hazard.

BUGSModelLibrary

http://bugsmodellibrary.googlecode.com

BUGSModelLibrary is a prototype PKPD model library for use with WinBUGS 1.4.3. The current version includes:

- Specific linear compartmental models:
  - One compartment model with first order absorption
  - Two compartment model with elimination from and first order absorption into central compartment
- General linear compartmental model described by a matrix exponential
- General compartmental model described by a system of first order ODEs

BUGSModelLibrary

The models and data format are based on NONMEM/NMTRAN/PREDPP conventions including:

- Recursive calculation of model predictions
  - This permits piecewise constant covariate values
- Bolus or constant rate inputs into any compartment
- Handles single dose, multiple dose and steady-state dosing histories
- Implemented NMTRAN data items include:
  - TIME, EVID, CMT, AMT, RATE, ADDL, II, SS

BUGSModelLibrary
Calling conventions

- All BUGSModelLibrary functions have the form:

  <modelName>`(time, amt, rate, ii, evid, cmt, addl, ss, theta)`

  where `time`, `amt`, `rate`, `ii`, `evid`, `cmt`, `addl` and `ss` are equal length time-ordered vectors and are defined identically to the NONMEM variables of the same name.
- `theta` may be either
  - A vector of parameters of length equal to the number of model parameters, or
  - A matrix with a number of rows equal to the length of the time vector. The i^{th} row contains a vector of model parameters for the time interval (`time[i-1]`, `time[i]`]. This permits time-dependent parameters.
- <modelName> is the name of a built-in model, such as `OneCptModel` or `TwoCptModel`, or a user-defined name.
- Within a WinBUGS model <modelName> is usually called once per individual within a `for` block that loops over individuals.

BUGSModelLibrary: Built-in models

- `OneCptModel(time, amt, rate, ii, evid, cmt, addl, ss, theta)`
- `TwoCptModel(time, amt, rate, ii, evid, cmt, addl, ss, theta)`
- `OneCptKaModel(time, amt, rate, ii, evid, cmt, addl, ss, theta)`
- `TwoCptKaModel(time, amt, rate, ii, evid, cmt, addl, ss, theta)`

model	WinBUGS function name	argument names	model parameters in `theta`
one compartment model with first order absorption $\left(k_a > \frac{CL}{V_2}\right)$	`OneCptModel`	`time, amt, rate, ii, evid, cmt, addl, ss`	CL, V_2, $k_a - \frac{CL}{V_2}$, F_1, F_2, t_{lag1}, t_{lag2}
two compartment model with first order absorption $(k_a > \lambda_1)$	`TwoCptModel`	`time, amt, rate, ii, evid, cmt, addl, ss`	CL, Q, V_2, V_3, $k_a - \lambda_1$, F_1, F_2, F_3, t_{lag1}, t_{lag2}, t_{lag3}
one compartment model with first order absorption	`OneCptKaModel`	`time, amt, rate, ii, evid, cmt, addl, ss`	CL, V_2, k_a, F_1, F_2, t_{lag1}, t_{lag2}
two compartment model with first order absorption	`TwoCptKaModel`	`time, amt, rate, ii, evid, cmt, addl, ss`	CL, Q, V_2, V_3, k_a, F_1, F_2, F_3, t_{lag1}, t_{lag2}, t_{lag3}

BUGSModelLibrary
User-programmed models

- Linear compartmental models
  - User specifies the non-zero elements of the rate constant matrix.
  - Linear ODE's are solved using matrix exponential methods.
- General compartmental models
  - User specifies the ODE's.
  - The ODE's are solved using either a Runge-Kutta or an adaptive multistep (LSODA) method
- Both cases require user specification of a rate constant matrix or ODE's in a template Component Pascal procedure that must be compiled using the BlackBox Component Builder 1.5.

Linear compartment model

- $\equiv$ model described by a system of first order linear differential equations with (piecewise) constant coefficients:

$$x'(t) = Kx(t)$$

  where K is a matrix.
- For example K for a two compartment PK model with first order absorption is:

$$K = \begin{bmatrix} -k_a & 0 & 0 \\ k_a & -(k_{10} + k_{12}) & k_{21} \\ 0 & k_{12} & -k_{21} \end{bmatrix}$$

- When applicable this method is usually faster than the Runge-Kutta or LSODA methods.

Linear compartment model

Component Pascal code for a two compartment PK model with first order absorption

```
PROCEDURE UserKMatrix(IN theta: ARRAY OF REAL;
                      nCmt: INTEGER):
    POINTER TO ARRAY OF ARRAY OF REAL;
  VAR
    kMatrix: POINTER TO ARRAY OF ARRAY OF REAL;
    i, j: INTEGER;
    k10, k12, k21, ka: REAL;
BEGIN
    NEW(kMatrix,nCmt,nCmt);
    (* Initialize to all zeros *)
    FOR i := 0 TO nCmt-1 DO;
    FOR j := 0 TO nCmt-1 DO;
      kMatrix[i,j] := 0;
    END; END;
    k10 := theta[0];
    k12 := theta[1];
    k21 := theta[2];
    ka := theta[3];
```

```
    (* Assign nonzero rate constants *)
    kMatrix[0,0] := -ka;
    kMatrix[1,0] := ka;
    kMatrix[1,1] := -(k10+k12);
    kMatrix[1,2] := k21;
    kMatrix[2,1] := k12;
    kMatrix[2,2] := -k21;

    RETURN kMatrix;

END UserKMatrix;

PROCEDURE (m: MatExpModel) InitModel*;
BEGIN
    m.nParameter := 10;
    m.F1Index := 4;
    m.tlag1Index := 7;
    m.nCmt := 3;
END InitModel;
```

Only the red portions need to be programmed by the user. The remainder is from a template provided with BUGSModelLibrary.

Demo: Two compartment PK model with effect compartment and first order absorption

See the BUGSModelLibrary User Manual, pp 11–13

General compartment model

- $\equiv$ model described by a system of first order ordinary differential equations (ODE's), i.e., differential equations of the form:

$$x'(t) = f(t, x(t))$$

  where x and f are vector-valued functions.
- For example a two compartment PK model with Michaelis-Menten elimination:

$$\begin{aligned} x_1'(t) &= -\left(\frac{V_{\max}}{K_m + x_1(t)} + k_{12}\right) x_1(t) + k_{21} x_2(t) \\ x_2'(t) &= k_{12} x_1(t) - k_{21} x_2(t) \end{aligned}$$

- Two ODE solving methods are available:
  - A Runge-Kutta 4th/5th order method—usually faster for non-stiff problems.
  - LSODA, the Livermore Solver for Ordinary Differential equations with Automatic method switching for stiff and nonstiff problems

General compartment model

Component Pascal code for a two compartment PK model with Michaelis-Menten elimination

```
PROCEDURE UserDerivatives(IN theta, x: ARRAY OF REAL;
numEq: INTEGER; t: REAL; OUT dxdt: ARRAY OF REAL) ;
VAR
    Vmax, Km, k12, k21: REAL;
BEGIN
    Vmax := theta[0];
    Km := theta[1];
    k12 := theta[2];
    k21 := theta[3];

    (* Differential equations for the model excluding piecewise *)
    (* constant input rates provided in the data set *)

    dxdt[0] := -(Vmax / (Km + x[0]) + k12) * x[0] + k21 * x[1];
    dxdt[1] := k12 * x[0] - k21 * x[1];

END UserDerivatives;
```


Demo: Two compartment PK model + indirect effect model with drug effect on kout (inhibitory Emax)

See the BUGSModelLibrary User Manual, pp 13–19

Specifying a non-standard sampling distribution in BUGS: The zeros trick

- Suppose the likelihood for the i^{th} observation has the general form $p(y_i|\theta, x_i)$ where p is a known probability density function that is not among the distributions built into WinBUGS.
- We can rewrite the likelihood as

$$p(y_i|\theta, x_i) = e^{\ell_i} = \frac{e^{-(-\ell_i)}(-\ell_i)^0}{0!} = p_{\text{Poisson}}(0|-\ell_i)$$

  where $\ell_i = \log(p(y_i|\theta, x_i))$, i.e., the log-likelihood, and p_{Poisson} is the Poisson distribution.
- So for each real observation we create a pseudo-random variable equal to 0 from a Poisson distribution with mean equal to minus the log-likelihood.
- To assure that the mean of the Poisson distribution is non-negative, add a positive constant to ℓ_i, i.e., use $p_{\text{Poisson}}(0|-\ell_i + C)$ where C is large enough to assure $-\ell_i + C > 0$. This does not affect the posterior distribution

Specifying a non-standard sampling distribution in BUGS: The zeros trick

```
model{

  for(i in 1:nObs){

    zeros[i] <- 0
    zeros[i] ~ dpois(phi[i])
    phi[i] <- -log(L[i]) + C
    L[i] <- ··· # user-specified likelihood function

  }

  C <- 10000

      ⋮
}
```

Specifying a non-standard sampling distribution in BUGS: The ones trick

- Alternatively we can rewrite the likelihood in terms a Bernoulli distribution:

$$p(y_i|\theta, x_i) = L_i = L_i^1 (1 - L_i)^0 = p_{\text{Bernoulli}}(1|L_i)$$

  where $L_i = (p(y_i|\theta, x_i))$, i.e., the likelihood, and $p_{\text{Bernoulli}}$ is the Bernoulli distribution.
- So for each real observation we create a pseudo-random variable equal to 1 from a Bernoulli distribution with mean equal to the likelihood.
- To assure that the probability parameter in the Bernoulli distribution is less than 1, divide L_i by a positive constant, i.e., use $p_{\text{Bernoulli}}(1|L_i/C)$ where C is large enough to assure $L_i/C \in (0, 1)$. This does not affect the posterior distribution

Specifying a non-standard sampling distribution in BUGS: The ones trick

```
model{

  for(i in 1:nObs){

    ones[i] <- 1
    ones[i] ~ dbern(p[i])
    p[i] <- L[i] / C
    L[i] <- ··· # user-specified likelihood function

  }

  C <- 10000

      ⋮
}
```

Example: PK/PD model of ondansetron anti-emetic effect

A Population Pharmacokinetic–Pharmacodynamic Analysis of Repeated Measures Time-to-Event Pharmacodynamic Responses: The Antiemetic Effect of Ondansetron*

Eugène H. Cox,[1] Christine Veyrat-Follet,[2] Stuart L. Beal,[3] Eliane Fuseau,[4] Saraswati Kenkare,[5] and Lewis B. Sheiner[3,6,7]

Journal of Pharmacokinetics and Biopharmaceutics 27: 625–644 (1999)

Example: PK/PD model of ondansetron anti-emetic effect

- NONMEM code and data available at `ftp://nonmem.iconplc.com` in /Public/nonmem/non_continuous/general_hazard
  - Data appear to be similar in form to that used for the publication but not the same—probably simulated data to illustrate the model.
- Subjects with count data were excluded—ran out of time to implement that complication.
- 63 healthy subjects
- Single 5 minute infusions of ondansetron (0.1, 0.25, 1, 4 or 8 mg)
- Ipecac administered at 0.5, 4, 6, 8, 12 or 16 h after ondansetron administration
- Times of emesis events are recorded.

Cox et al model

Time of the i^{th} emesis event relative to the previous emesis event or ipecac dose in the j^{th} individual:

$$
\begin{aligned}
\Delta t_{\text{emesis},ij} &= t_{\text{emesis},ij} - t_{\text{prev},ij} \sim p_{\text{emesis}}\left(\theta_{\text{PD},j}, \theta_{\text{PK},j}, D_j, t_{\text{prev},ij}, t_{\text{ipecac},j}\right) \\
t_{\text{prev},ij} &= \max\left(t_{\text{emesis},(i-1)j}, t_{\text{ipecac},j}\right) \\
\theta_{\text{PD},j} &= \left(\theta_{1j}, \theta_2, \theta_3, \theta_4\right) \\
\theta_{\text{PK},j} &= \left(CL_j, Q_j, V1_j, V2_j\right) \\
p_{\text{emesis}}&\left(\Delta t_{\text{emesis},ij} | \theta_{\text{PD},j}, \theta_{\text{PK},j}, D_j, t_{\text{prev},ij}, t_{\text{ipecac},j}\right) \\
&= h\left(\Delta t_{\text{emesis},ij} + t_{\text{prev},ij} | \theta_{\text{PD},j}, \theta_{\text{PK},j}, D_j, t_{\text{ipecac},j}\right) \\
&\quad \times e^{-\int_{t_{\text{prev},ij}}^{\Delta t_{\text{emesis},ij}+t_{\text{prev},ij}} h\left(u | \theta_{\text{PD},j}, \theta_{\text{PK},j}, D_j, t_{\text{ipecac},j}\right) du} \\
\log&\left(h\left(t | \theta_{\text{PD},j}, \theta_{\text{PK},j}, D_j, t_{\text{ipecac},j}\right)\right) \\
&= \theta_{1j} + \theta_2 \log\left(t - t_{\text{ipecac},j}\right) - \theta_3\left(t - t_{\text{ipecac},j}\right) \\
&\quad -\theta_4 C_{\text{ond},j}\left(t, D_j, \theta_{\text{PK},j}\right) \\
\theta_{1j} &\sim N\left(\widehat{\theta}_1, \omega^2_{\theta_1}\right)
\end{aligned}
$$

Course project

One of the requirements for course credit is completion of a small population modeling project. This project represents 50% of the course grade. The project is due within two weeks of the final course lecture. Please direct any questions to course instructor.

Project requirements

Students will define their own projects, according to the following required elements:

- Choose a real-world or simulated data set suitable for modeling with a Bayesian hierarchical model for categorical, count or time-to-event data (e.g., a population PD model).
- Develop and implement a suitable model using WinBUGS. Use R for data management, launching WinBUGS and analysis of the MCMC simulation results.
- Summarize methods, results and conclusions with supporting figures and tables in a brief report. Report format should be PDF document, if possible. The results should include suitable model evaluation plots, e.g., posterior predictive checks.
- Discuss results including assumptions inherent in the analysis. This should include rationale for the prior distribution.
- Include supplemental files:
  - Data file(s)
  - R script(s) and WinBUGS model file(s)

Selected references

Bayesian modeling, statistics and decision analysis


A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin.
Bayesian data analysis.
Chapman & Hall/CRC, New York, 2004.


B.P. Carlin and T.A. Louis.
Bayes and Empirical Bayes Methods for Data Analysis.
Chapman & Hall/CRC Press, 2nd edition, 2000.


J.O. Berger.
Statistical Decision Theory and Bayesian Analysis.
Springer, 2nd edition, 1993.


D. J. Lunn, J. Wakefield, and A. Racine-Poon.
Cumulative logit models for ordinal data: a case study involving allergic rhinitis severity scores.
Stat Med, 20(15):2261–85, 2001.


D.J. Spiegelhalter, K.R. Abrams, and J.P. Myles.
Bayesian Approaches to Clinical Trials and Health-Care Evaluation.
Wiley, 2004.

Selected references I
Modeling of categorical, count and time-to-event data

A Agresti.
Categorical Data Analysis.
Wiley, 2nd edition, 2002.

J.O. Berger.
Statistical Decision Theory and Bayesian Analysis.
Springer, 2nd edition, 1993.

P Congdon.
Bayesian Models for Categorical Data.
Wiley, 2005.

E. H. Cox, C. Veyrat-Follet, S. L. Beal, E. Fuseau, S. Kenkare, and L. B. Sheiner.
A population pharmacokinetic-pharmacodynamic analysis of repeated measures time-to-event pharmacodynamic responses: the antiemetic effect of ondansetron.
Journal of Pharmacokinetics and Biopharmaceutics, 27(6):625–644, 1999.

E. I. Ette, A. Roy, and P. Nandy.
Population pharmacokinetic/pharmacodynamic modeling of ordered categorical longitudinal data.
In E.I. Ette and P.J. Williams, editors, *The Science of Quantitative Pharmacology*, pages 655–688. Wiley, 2007.

Selected references II
Modeling of categorical, count and time-to-event data

J Fiedler-Kelly.
Pk/pd analysis of binary (logistic) outcome data.
In E.I. Ette and P.J. Williams, editors, *The Science of Quantitative Pharmacology*, pages 633–654. Wiley, 2007.

CJ Godfrey.
Mixed effects modeling analysis of count data.
In E.I. Ette and P.J. Williams, editors, *The Science of Quantitative Pharmacology*, pages 699–721. Wiley, 2007.

J.P. Klein and M.L. Moeschberger.
Survival Analysis. Techniques for Censored and Truncated Data.
Springer, 1997.

L. B. Sheiner, S. L. Beal, and A. Dunne.
Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials.
Journal of the American Statistical Association, 92(440):1235–1244, 1997.

T. M. Therneau and P. M. Grambsch.
Modeling survival data: Extending the cox model, 2000.

COMPLETE COURSEWARE LICENSE AGREEMENT I

RIGHTS RESERVED: THIS COMPLETE COURSEWARE LICENSE IS SEPARATE AND DISTINCT FROM THE OPEN COURSEWARE LICENSE, WHICH IS PROVIDED FOR PUBLIC USE OF A SPECIFIED SUBSET OF METRUM INSTITUTE TRAINING MATERIALS. SEE “Courseware Info” at http://metruminstitute.org.

THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED. BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. TO THE EXTENT THIS LICENSE MAY BE CONSIDERED TO BE A CONTRACT, THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

1. Definitions
“Adaptation” means a work based upon the Work, or upon the Work and other pre-existing works, such as a translation, adaptation, derivative work, or any other form in which the Work may be recast, transformed, or adapted including in any form recognizably derived from the original, except that a work that constitutes a Collection will not be considered an Adaptation for the purpose of this License.
“Collection” means a collection of literary or artistic works, such as encyclopedias and anthologies, or performances, phonograms or broadcasts, or other works or subject matter other than works listed in Section 1(f) below, which, by reason of the selection and arrangement of their contents, constitute intellectual creations, in which the Work is included in its entirety in unmodified form along with one or more other contributions, each constituting separate and independent works in themselves, which together are assembled into a collective whole. A work that constitutes a Collection will not be considered an Adaptation (as defined above) for the purposes of this License.
“Distribute” means to make available to anyone other than You, the original and copies of the Work or Adaptation, as appropriate, through sale or other transfer of ownership.
“Licensor” means the individual, individuals, entity or entities that offer(s) the Work under the terms of this License.
“Original Author” means the individual, individuals, entity or entities who created the Work or if no individual or entity can be identified, the publisher.
“Work” means the scientific work offered under the terms of this License including without limitation any production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression including digital form, such as a book, pamphlet and other writing; a lecture, address, sermon or other work of the same nature; a broadcast; a phonogram; a compilation of data to the extent it is protected as a copyrightable work.

COMPLETE COURSEWARE LICENSE AGREEMENT II

“You” means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.
“Publicly Perform” means to perform public recitations of the Work and to communicate to the public those public recitations, by any means or process, including by wire or wireless means or public digital performances; to make available to the public Works in such a way that members of the public may access these Works from a place and at a place individually chosen by them; to perform the Work to the public by any means or process and the communication to the public of the performances of the Work, including by public digital performance; to broadcast and rebroadcast the Work by any means
including signs, sounds or images.
“Reproduce” means to make copies of the Work by any means including without limitation by sound or visual recordings and the right of fixation and reproducing fixations of the Work, including storage of a protected performance or phonogram in digital form or other electronic medium.
“Commercial” means any copy, application or Adaptation of the Work for the direct or indirect purpose of revenue generation
“Non-Commercial” means any copy, application or Adaptation of the Work for purposes that are not directly or indirectly related to revenue generation.

2. Fair Dealing Rights. Nothing in this License is intended to reduce, limit, or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws.

3. License Grant. Subject to the terms and conditions of this License, Licensor hereby grants You fair use of this copyrighted Work for *noncommercial educational* purposes or for commercial and noncommercial *research* purposes. For example, You may adapt example code sets (the Work) and apply them to a commercial research project.

4. Restrictions. The license granted in Section 3 above is expressly made subject to and limited by the following restrictions: Under the terms of this License, You may not Distribute, Reproduce, or Publicly Perform the Work. For example, You may not use courseware (the Work or any adaptations) to create a new public training course, whether commercial or non-commercial. You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties.

COMPLETE COURSEWARE LICENSE AGREEMENT III

For the avoidance of doubt:
Non-waivable Compulsory License Schemes. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme cannot be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License;
Waivable Compulsory License Schemes. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme can be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License if Your exercise of such rights is for a purpose or use which is otherwise than noncommercial as permitted under Section 4(b) and otherwise waives the right to collect royalties through any statutory or compulsory licensing scheme; and,
Voluntary License Schemes. The Licensor reserves the right to collect royalties, whether individually or, in the event that the Licensor is a member of a collecting society that administers voluntary licensing schemes, via that society, from any exercise by You of the rights granted under this License that is for a purpose or use which is otherwise than noncommercial as permitted under Section 4(c).

5. Representations, Warranties and Disclaimer
UNLESS OTHERWISE MUTUALLY AGREED TO BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

6. Limitation on Liability. EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. Termination This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms

COMPLETE COURSEWARE LICENSE AGREEMENT IV

or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

8. Miscellaneous If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You. The rights granted under, and the subject matter referenced, in this License were drafted utilizing the terminology of the Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), the Rome Convention of 1961, the WIPO Copyright Treaty of 1996, the WIPO Performances and Phonograms Treaty of 1996 and the Universal Copyright Convention (as revised on July 24, 1971). These rights and subject matter take effect in the relevant jurisdiction in which the License terms are sought to be enforced according to the corresponding provisions of the implementation of those treaty provisions in the applicable national law. If the standard suite of rights granted under applicable copyright law includes additional rights not granted under this License, such additional rights are deemed to be included in the License; this License is not intended to restrict the license of any rights under applicable law.