2
Motivating Studies

2.1 Introduction

In this chapter, we present a number of studies that motivate this work and/or are used repeatedly throughout the text. Upon going through the book, the reader will find more examples. These are either used once or at least confined to one or a few chapters. A single-arm clinical trial conducted in patients with chronic pain, the analgesic trial, is introduced in Section 2.2. Section 2.3 is devoted to a two-armed clinical trial in patients treated for toenail infection. The fluvoxamine study, a post-marketing study conducted in psychiatric patients, is introduced in Section 2.4. A controlled clinical trial, conducted in patients suffering from epileptic seizures, is presented in Section 2.5. All studies introduced thus far are longitudinal in nature. Section 2.6 discusses the Project on Preterm and Small for Gestational Age Infants study (POPS), an epidemiologic study in which interest lies in a multivariate outcome. A key clustered data example from the developmental toxicology area, conducted under the U.S. National Toxicology Program (NTP), is presented in Section 2.7. Section 2.8 introduces the sports injuries trial, studying two longitudinal post-operative outcomes. Finally, Section 2.9 is devoted to the Age Related Macular Degeneration Study (ARMD), an ophthalmologic clinical trial in which both a continuous as well as a categorical longitudinally measured outcome is of interest.

2.2 The Analgesic Trial

These data come from a single-arm clinical trial in 395 patients who are given analgesic treatment for pain caused by chronic nonmalignant disease. Treatment was to be administered for 12 months and assessed by means of a 'Global Satisfaction Assessment' (GSA) scale, rated on a five-point scale:

$$\text{GSA} = \begin{cases} 1 & : \text{ very good,} \\ 2 & : \text{ good,} \\ 3 & : \text{ indifferent,} \\ 4 & : \text{ bad,} \\ 5 & : \text{ very bad.} \end{cases} \tag{2.1}$$

Many of our analyses will focus on a dichotomized version, defined in (17.1), but Chapter 18 will consider the ordinal version of the outcome. Apart from the outcome of interest, a number of covariates are available, such as age, sex, weight, duration of pain in years prior to the start of the study, type of pain, physical functioning, psychiatric condition, respiratory problems, etc.

GSA was rated by each person four times during the trial, at months 3, 6, 9, and 12. An overview of the frequencies per follow up time is given in Table 2.1. Inspecting Table 2.1 reveals that the total per column is variable. This is due to missingness. At three months, 10 subjects lack a measure, with these numbers being 93, 168, and 172 at subsequent times. Not only monotone missingness or dropout occurs, there are also subjects with intermittent values.

An overview of the extent of missingness is shown in Table 2.2. Note that only around 40% of the subjects have complete data. The dropout sequences amount to roughly another 40%, with close to 20% of the patterns showing intermittent missingness. This example underscores that a satisfactory longitudinal analysis will oftentimes have to address the missing data problem.

2.3 The Toenail Data

The data introduced in this section were obtained from a randomized, double-blind, parallel group, multicenter study for the comparison of two oral treatments (in what follows coded as A and B) for toenail dermatophyte onychomycosis (TDO), described in full detail by De Backer *et al* (1996). TDO is a common toenail infection, difficult to treat, affecting more than 2 out of 100 persons (Roberts 1992). Antifungal compounds, classically used for treatment of TDO, need to be taken until the whole nail has grown out healthy. The development of new compounds, however,

TABLE 2.1. *Analgesic Trial. Absolute and relative frequencies of the five GSA categories for each of the four follow up times.*

GSA	Month 3		Month 6		Month 9		Month 12	
1	55	14.3%	38	12.6%	40	17.6%	30	13.5%
2	112	29.1%	84	27.8%	67	29.5%	66	29.6%
3	151	39.2%	115	38.1%	76	33.5%	97	43.5%
4	52	13.5%	51	16.9%	33	14.5%	27	12.1%
5	15	3.9%	14	4.6%	11	4.9%	3	1.4%
Tot	385		302		227		223	

TABLE 2.2. *Analgesic Trial. Overview of missingness patterns and the frequencies with which they occur. 'O' indicates observed and 'M' indicates missing.*

Measurement occasion					
Month 3	Month 6	Month 9	Month 12	Number	%
Completers					
O	O	O	O	163	41.2
Dropouts					
O	O	O	M	51	12.91
O	O	M	M	51	12.91
O	M	M	M	63	15.95
Non-monotone missingness					
O	O	M	O	30	7.59
O	M	O	O	7	1.77
O	M	O	M	2	0.51
O	M	M	O	18	4.56
M	O	O	O	2	0.51
M	O	O	M	1	0.25
M	O	M	O	1	0.25
M	O	M	M	3	0.76

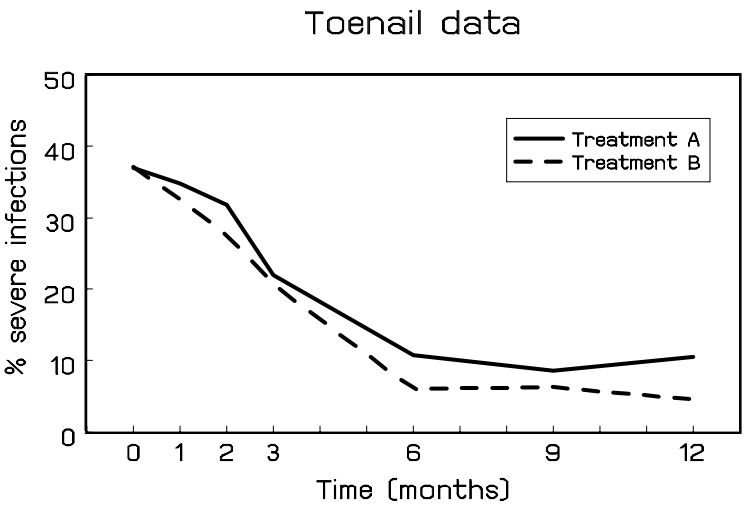

FIGURE 2.1. *Toenail Data. Evolution of the percentage of severe toenail infections in the two treatment groups separately.*

has reduced the treatment duration to 3 months. The aim of the present study was to compare the efficacy and safety of 12 weeks of continuous therapy with treatment A or with treatment B.

In total, 2×189 patients were randomized, distributed over 36 centers. Subjects were followed during 12 weeks (3 months) of treatment and followed further, up to a total of 48 weeks (12 months). Measurements were taken at baseline, every month during treatment, and every 3 months afterwards, resulting in a maximum of 7 measurements per subject. At the first occasion, the treating physician indicates one of the affected toenails as the target nail, the nail which will be followed over time. We will restrict our analyses to only those patients for which the target nail was one of the two big toenails. This reduces our sample under consideration to 146 and 148 subjects, in group A and group B, respectively.

One of the responses of interest was the unaffected nail length, measured from the nail bed to the infected part of the nail, which is always at the free end of the nail, expressed in mm. This outcome has been studied extensively in Verbeke and Molenberghs (2000). Another important outcome in this study was the severity of the infection, coded as 0 (not severe) or 1 (severe). The question of interest was whether the percentage of severe infections decreased over time, and whether that evolution was different for the two treatment groups. A summary of the number of patients in the study at each time-point, and the number of patients with severe infections is given in Table 2.3. A graphical representation is given in Figure 2.1.

Due to a variety of reasons, the outcome has been measured at all 7 scheduled time points, for only 224 (76%) out of the 298 participants. Table 2.4 summarizes the number of available repeated measurements per

TABLE 2.3. *Toenail Data. Number and percentage of patients (N) with severe toenail infection, for each treatment arm separately.*

	Group A			Group B		
	# Severe	*N*	%	# Severe	*N*	%
Baseline	54	146	37.0%	55	148	37.2%
1 month	49	141	34.7%	48	147	32.6%
2 months	44	138	31.9%	40	145	27.6%
3 months	29	132	22.0%	29	140	20.7%
6 months	14	130	10.8%	8	133	6.0%
9 months	10	117	8.5%	8	127	6.3%
12 months	14	133	10.5%	6	131	4.6%

TABLE 2.4. *Toenail Data. Number of available repeated measurements per subject, for each treatment arm separately.*

	Group A		Group B	
# Obs.	*N*	%	*N*	%
1	4	2.74%	1	0.68%
2	2	1.37%	1	0.68%
3	4	2.74%	3	2.03%
4	2	1.37%	4	2.70%
5	2	1.37%	8	5.41%
6	25	17.12%	14	9.46%
7	107	73.29%	117	79.05%
Total:	146	100%	148	100%

subject, for both treatment groups separately. We see that the occurrence of missingness is similar in both treatment groups.

2.4 The Fluvoxamine Trial

Accumulated experience with fluvoxamine, a serotonin reuptake inhibitor, in controlled clinical trials has shown it to be as effective as conventional antidepressant drugs and more effective than placebo in the treatment of depression (Burton 1991). However, many patients who suffer from depression have concomitant morbidity with conditions such as obsessive-compulsive disorder, anxiety disorders and, to some extent, panic disorders. In most trials, patients with comorbidity are excluded, and therefore, it is of interest to gather evidence as to the importance of such factors, with a view on improved diagnosis and treatment. The general aim of this study was to determine the profile of fluvoxamine in ambulatory clinical psychiatric practice.

A total of 315 patients were enrolled with one or more of the following diagnoses: depression, obsessive, compulsive disorder, and panic disorder. Several covariates were recorded, such as gender and initial severity on a 5-point ordinal scale, where severity increases with category. After recruitment of the patient in the study, he or she was investigated at four visits (weeks 2, 4, 8, and 12). On the basis of about twenty psychiatric symptoms, the therapeutic effect and the side-effects were scored at each visit in an ordinal manner. Side effect is coded as (1) = no; (2) = not interfering with functionality of patient; (3) = interfering significantly with functionality of patient; (4) = the side-effect surpasses the therapeutic effect. Similarly, the effect of therapy is recorded on a four-point ordinal scale: (1) no improvement over baseline or worsening; (2) minimal improvement (not changing functionality); (3) moderate improvement (partial disappearance of symptoms); and (4) important improvement (almost disappearance of symptoms). Thus, a side effect occurs if new symptoms occur while there is therapeutic effect if old symptoms disappear. These data were used, among others, by Molenberghs and Lesaffre (1994), Molenberghs, Kenward, and Lesaffre (1997), and Lapp, Molenberghs, and Lesaffre (1998), Van Steen *et al* (2001), and Jansen *et al* (2003).

Table 2.5 gives the absolute and relative frequencies over the four categories of side effects and therapeutic effect for each of the four follow-up times. Because there are 315 subjects enrolled in the trial, it is clear that at the four times there are 16, 46, 72, and 89 subjects missing, respectively. The missing data patterns, common to both outcomes, are represented in Table 2.6. Note that a much larger fraction is fully observed than in, for example, the analgesic trial (Section 2.2). Among the incomplete sequences, dropout is much more common than intermittent missingness, the latter

TABLE 2.5. *Fluvoxamine Trial. Absolute and relative frequencies of the four side effects and therapeutic effect categories for each of the four follow-up times.*

	Week 2		Week 4		Week 8		Week 12	
Side effects								
0	128	42.8%	144	52.5%	156	64.2%	148	65.5%
1	128	42.8%	103	38.3%	79	32.5%	71	31.4%
2	28	9.4%	17	6.3%	6	2.5%	7	3.1%
3	15	5.2%	5	1.9%	2	0.8%	0	0.0%
Therapeutic effects								
0	19	6.4%	64	23.8%	110	45.3%	135	59.7%
1	95	31.8%	114	42.4%	93	38.3%	62	27.4%
2	102	34.1%	62	23.1%	30	12.4%	19	8.4%
3	83	27.8%	29	10.8%	10	4.1%	10	4.4%
Tot	299		269		243		226	

TABLE 2.6. *Fluvoxamine Trial. Overview of missingness patterns and the frequencies with which they occur. 'O' indicates observed and 'M' indicates missing.*

Measurement occasion					
Month 3	Month 6	Month 9	Month 12	Number	%
Completers					
O	O	O	O	224	71.11
Dropouts					
O	O	O	M	18	5.71
O	O	M	M	26	8.25
O	M	M	M	31	9.84
M	M	M	M	14	4.44
Non-monotone missingness					
M	O	O	O	1	0.32
M	M	M	O	1	0.32

type confined to two sequences only. Observe that, unlike in Table 2.2, there are subjects, 14 in total without any follow-up measurements. This group of subjects is still an integral part of the trial, as they contain baseline information, including covariate information and baseline assessment of severity of the mental illness.

2.5 The Epilepsy Data

The data considered here are obtained from a randomized, double-blind, parallel group multi-center study for the comparison of placebo with a new anti-epileptic drug (AED), in combination with one or two other AED's. The study is described in full detail in Faught *et al* (1996). The randomization of epilepsy patients took place after a 12-week baseline period that served as a stabilization period for the use of AED's, and during which the number of seizures were counted. After that period, 45 patients were assigned to the placebo group, 44 to the active (new) treatment group. Patients were then measured weekly. Patients were followed (double-blind) during 16 weeks, after which they were entered into a long-term open-extension study. Some patients were followed for up to 27 weeks. The outcome of interest is the number of epileptic seizures experienced during the last week, i.e., since the last time the outcome was measured. The key research question is whether or not the additional new treatment reduces the number of epileptic seizures. As a summary of the data, Figure 2.2 shows a frequency plot, over all visits, over both treatment groups. We observe a very skewed distribution, with largest observed value equal to 73 seizures in one week time. Average and median evolutions are shown in Figure 2.3. The unstable behavior can be explained by the presence of extreme values, but is also the result of the fact that very little observations are available at some of the time-points, especially past week 20. This is also reflected in Table 2.7, which shows the number of measurements at a selection of time-points. Note the serious drop in number of measurements past the end of the actual double-blind period, i.e., past week 16.

2.6 The Project on Preterm and Small for Gestational Age Infants (POPS) Study

The Project On Preterm and Small-for-gestational age infants (POPS) collected information on 1338 infants born in The Netherlands in 1983 and having gestational age less than 32 weeks and/or birthweight less than 1500 g (Verloove *et al* 1988). In total, 133 clinics were involved. The study population represents 94% of the births in that year with similar gestational age and birthweight characteristics. Prenatal, perinatal, and postnatal in-

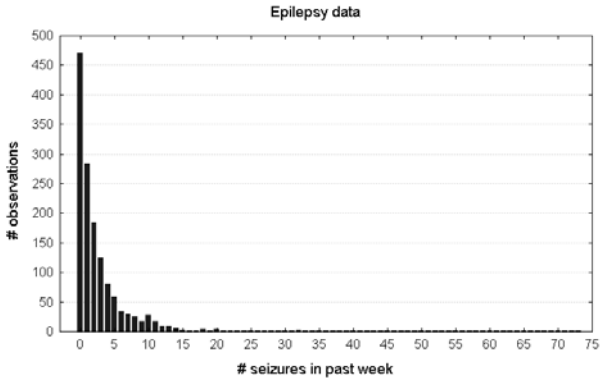

FIGURE 2.2. *Epilepsy Data. Frequency plot, over all visits, over both treatment groups.*

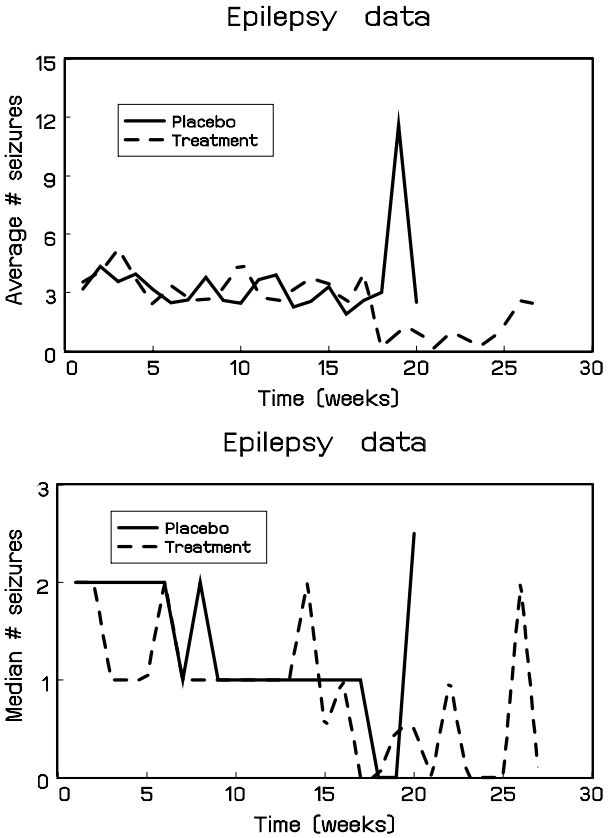

FIGURE 2.3. *Epilepsy Data. Average and median evolutions over time.*

TABLE 2.7. *Epilepsy Data. Number of measurements available at a selection of time-points, for both treatment groups separately.*

	# Observations		
Week	Placebo	Treatment	Total
1	45	44	89
5	42	42	84
10	41	40	81
15	40	38	78
16	40	37	77
17	18	17	35
20	2	8	10
27	0	3	3

formation as well as two year follow-up data were collected. Furthermore, the data base contains information on the delivery and specific details of the infant. After two years the child was reexamined. Lesaffre and Molenberghs (1991) and Molenberghs and Lesaffre (1994) studied the relationship between three ability scores measured at the age of two and risk factors measured at delivery. All ability scores were recorded in a dichotomous manner. They were available for 799 children. The first score ($ABIL_1$) checks whether the child can pile three bricks, $ABIL_1 = 1$ corresponds to 'no,' whereas $ABIL_1 = 2$ to 'yes.' The second score ($ABIL_2$) measures whether the physical movements of the child are natural, $ABIL_2 = 1$(no) and $ABIL_2 = 2$(yes). Although $ABIL_2$ is a purely physical ability score, $ABIL_1$ is a combination of physical and mental qualities. The third ability score, $ABIL_3$, expresses whether or not the child is able to put a ball in a box if he or she is asked to do so. The problem is to determine the risk factors for low performance at the three tests. Further it is of interest to compare the predicted probabilities taking into account the relationship between the responses to those calculated under the assumption of independent responses.

The defining variables of the POPS study, birth weight and gestational age, are shown graphically in Figure 2.4. It is clear from the figure that at least one of these needs to be low to be enrolled into the study.

The three ability scores are tabulated in Table 2.8. Of the 1338 subjects, 818 (61.1%) have all three ability scores observed, and 471 (35.2%) have none of them. Only 49 (3.7%) have partial information. The latter is not unexpected, since two years lapsed between enrollment and the assessment of the ability scores.

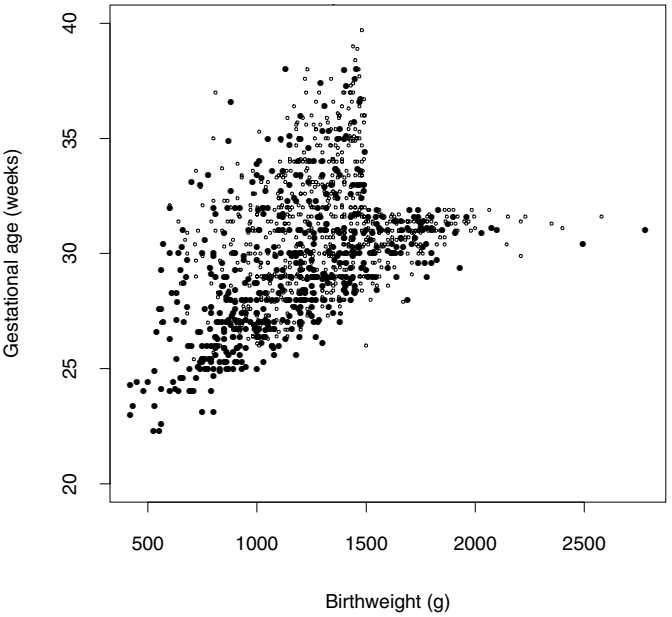

FIGURE 2.4. *POPS Study. The open circles correspond to zero outcomes.*

TABLE 2.8. *POPS Study. Frequency table of the three binary ability scores. Missing values are represented by M.*

		$ABIL_3$		
$ABIL_1$	$ABIL_2$	0	1	M
0	0	685	20	19
0	1	30	5	2
0	M	1	0	1
0	0	25	11	1
0	1	10	32	0
0	M	0	0	0
0	0	5	0	15
0	1	0	2	2
0	M	1	0	471

2.7 National Toxicology Program Data

The developmental toxicity studies introduced in this section are conducted at the Research Triangle Institute, which is under contract to the National

Toxicology Program of the United States (NTP data). These studies investigate the effects in mice of five chemicals: ethylene glycol (Price *et al* 1985), diethylene glycol dimethyl ether (Price *et al* 1987), and di(2--ethylhexyl)phthalate (Tyl *et al* 1988).

2.7.1 Ethylene Glycol

Ethylene glycol (EG) is also called 1,2-ethanediol and can be represented by the chemical formula $HOCH_2CH_2OH$. It is a high-volume industrial chemical with many applications. EG is used as an antifreeze in cooling and heating systems, as one of the components of hydraulic brake fluids, as an ingredient of electrolytic condensers, and as a solvent in the paint and plastics industries. Furthermore, EG is employed in the formulation of several types of inks, as a softening agent for cellophane, and as a stabilizer for soybean foam used to extinguish oil and gasoline fires. Also, one uses EG in the synthesis of various chemical products, such as plasticizers, synthetic fibers, and waxes (Windholz 1983).

EG may represent little hazard to human health in normal industrial handling, except possibly when used as an aerosol or at elevated temperatures. EG at ambient temperatures has a low vapor pressure and is not very irritating to the eyes or skin. However, accidental or intentional ingestion of antifreeze products, of which approximately 95% is EG, is toxic and may result in death (Rowe 1963, Price *et al* 1985).

Price *et al* (1985) describe a study in which timed-pregnant CD-1 mice were dosed by gavage with EG in distilled water. Dosing occurred during the period of organogenesis and structural development of the foetuses (gestational days 8 through 15). The doses selected for the study were 0, 750, 1500, or 3000 mg/kg/day. Table 2.9 shows, for each dose group and for all five NTP toxic agents, the number of dams containing at least one implant, the number of dams having at least one viable fetus, the number of live foetuses, the mean litter size, and the percentage of malformation for three different classes: external malformations, visceral malformations, and skeletal malformations. While for EG, skeletal malformations are substantial in the highest dose group, external and visceral malformations show only slight dose effects. The distribution of the number of implants is given in Table 2.10 for each of these five chemicals.

Figures 2.5 and 2.6 show for each of these studies and for each dose group the observed and averaged malformation rates in mice.

2.7.2 Di(2-ethylhexyl)Phthalate

Di(2-ethylhexyl)phthalate (DEHP) is also called octoil, dioctyl phthalate, or 1,2-benzenedicarboxylic acid bis(2-ethylhexyl) ester. It can be represented by $C_{24}H_{38}O_4$. DEHP is used in vacuum pumps (Windholz 1983).

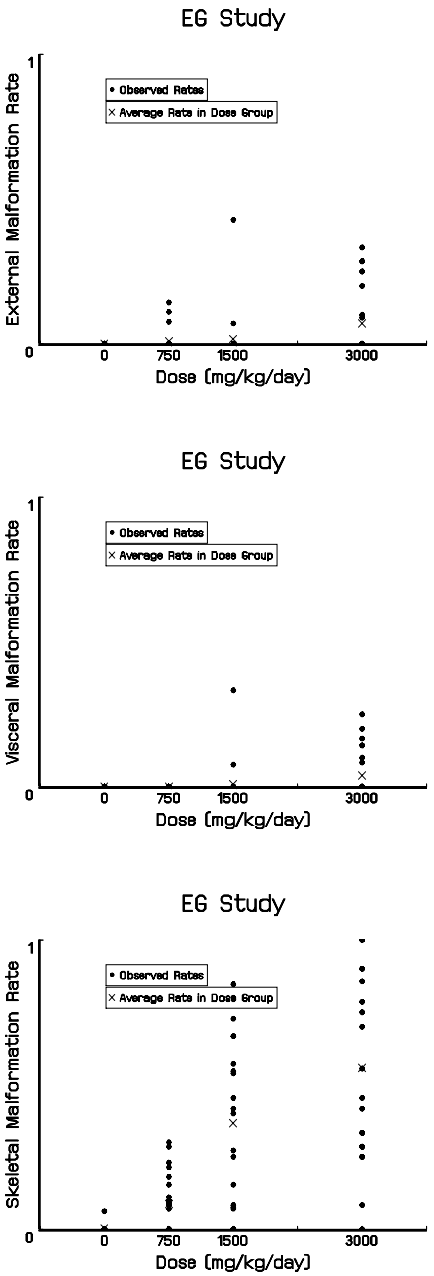

FIGURE 2.5. *NTP Data. EG Study in Mice. Observed and averaged malformation rates.*

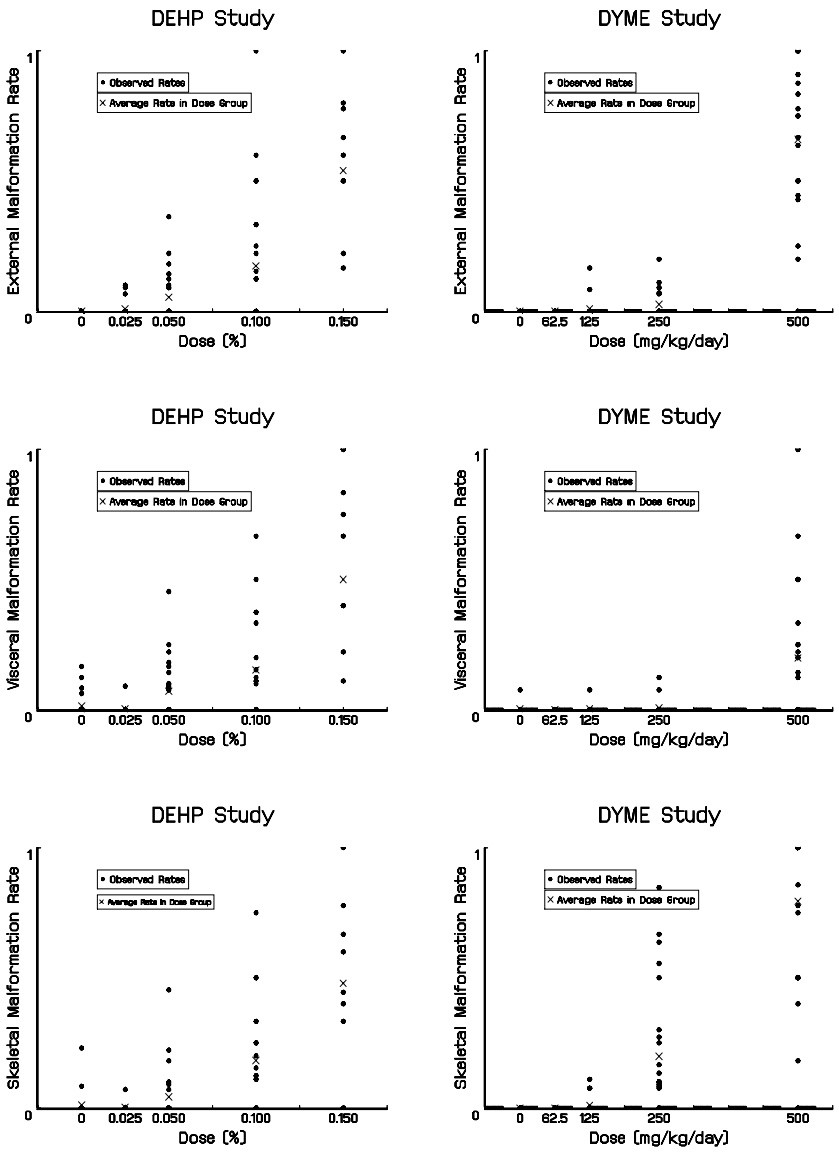

FIGURE 2.6. *NTP Data. DEHP and DYME Studies. Observed and averaged malformation rates.*

TABLE 2.9. *NTP Data. Summary data by study in mice. The dose is in mg/kg/day.*

		# Dams, ≥ 1			Litter Size	Malformations		
Exposure	Dose	Impl.	Viab.	Live	(mean)	Ext.	Visc.	Skel.
EG	0	25	25	297	11.9	0.0	0.0	0.3
	750	24	24	276	11.5	1.1	0.0	8.7
	1500	23	22	229	10.4	1.7	0.9	36.7
	3000	23	23	226	9.8	7.1	4.0	55.8
DEHP	0	30	30	330	13.2	0.0	1.5	1.2
	44	26	26	288	11.1	1.0	0.4	0.4
	91	26	26	277	10.7	5.4	7.2	4.3
	191	24	17	137	8.1	17.5	15.3	18.3
	292	25	9	50	5.6	54.0	50.0	48.0
DYME	0	21	21	282	13.4	0.0	0.0	0.0
	62.5	20	20	225	11.3	0.0	0.0	0.0
	125	24	24	290	12.1	1.0	0.0	1.0
	250	23	23	261	11.3	2.7	0.1	20.0
	500	22	22	141	6.1	66.0	19.9	79.4

Furthermore, this ester as well as other phthalic acid esters are used extensively as plasticizers for numerous plastic devices made of polyvinyl chloride. DEHP provides the finished plastic products with desirable flexibility and clarity (Shiota, Chou, and Nishimura 1980).

It has been well documented that small quantities of phthalic acid esters may leak out of polyvinyl chloride plastic containers in the presence of food, milk, blood, or various solvents. Due to their ubiquitous distribution and presence in human and animal tissues, considerable concern has developed as to the possible toxic effects of the phthalic acid esters (e.g., Autian 1973).

In particular, the developmental toxicity study described by Tyl *et al* (1988) has attracted much interest in the toxicity of DEHP. The doses selected for the study were 0, 0.025, 0.05, 0.1, and 0.15%, corresponding to a DEHP consumption of 0, 44, 91, 191, and 292 mg/kg/day, respectively. Females were observed daily during treatment, but no maternal deaths or distinctive clinical signs were observed. The dams were sacrificed, slightly prior to normal delivery, and the status of uterine implantation sites recorded. A total of 1082 live foetuses were dissected from the uterus, anesthetized, and examined for external, visceral, and skeletal malformations.

TABLE 2.10. *NTP Data. Frequency distribution of the number of implants.*

Number of implants	EG	DEHP	DYME
1	0	1	0
2	0	1	0
3	1	0	1
4	0	2	1
5	1	0	0
6	0	2	0
7	2	0	2
8	1	4	2
9	8	5	2
10	4	7	7
11	8	18	10
12	19	21	15
13	16	26	27
14	11	21	19
15	16	10	9
16	6	8	10
17	1	2	5
18	0	2	0
19	1	1	0
	95	131	110

Table 2.9 suggests clear dose-related trends in the malformation rates. The average litter size (number of viable animals) decreases with increased levels of exposure to DEHP, a finding that is attributable to the dose-related increase in fetal deaths.

2.7.3 Diethylene Glycol Dimethyl Ether

Other names for diethylene glycol dimethyl ether (DYME) are diglyme and bis(2-methoxyethyl) ether. DYME has as its chemical formula

$$CH_3O(CH_2)_2O(CH_2)_2OCH_3$$

(Windholz 1983). It is a component of industrial solvents. These are widely used in the manufacture of protective coatings such as lacquers, metal coatings, baking enamels, etc. (NIOSH 1983). Although to date, several at-

TABLE 2.11. *Sports Injuries Trial. Cross-classification of awakeness measurements at 10 and 20 minutes, on a four-point ordinal scale.*

Score at	Score at 20 mins			
10 mins	0	1	2	3
0	42	119	6	0
1	0	68	31	3
2	0	0	3	2
3	0	0	0	2

tempts have proven inadequate to evaluate the potential of glycol ethers to produce human reproductive toxicity, structurally related compounds have been identified as reproductive toxicants in several mammalian species, producing (1) testicular toxicity and (2) embryotoxicity.

Price *et al* (1987) describe a study in which timed-pregnant mice were dosed with DYME throughout major organogenesis (gestational days 8 through 15). The doses selected for the study were 0, 62.5, 125, 250 and 500 mg/kg/day. Table 2.9 summarizes the data.

2.8 The Sports Injuries Trial

These data come from a randomized, parallel group, double-blind trial in men comparing the effect of an active treatment to placebo on post-operative shivering and per-operative hemodynamics. The primary responses of interest were severity of post-operative shivering measured from the end of anesthesia every 5 minutes during 30 minutes as none (0), mild (1), moderate (2), or severe (3), and effect of treatment on overall consciousness assessed from the end of anesthesia at 10, 20, 30, 45, 60, 90, and 120 minutes as impossible to awake (0), difficult to awake (1), easy to awake (2), and awake, eyes open (3). One hundred forty patients were assigned to each treatment group.

Since this trial occurred in a very short time period, there is very little missing data. There was one patient who had no response information for either variable, so this patient is excluded from all analyses. There were also 3 patients with some missing information on shivering or overall consciousness, leaving 138 patients with complete information.

One interesting feature of these data is that there are structural zeros in the awake variables. A patient could never become less awake over time, thus the cross-tabulation of the score over time contains zeros in the lower left corner. Data from 10 and 20 minutes are presented in Table 2.11. The zero in the upper right hand corner (0 at 10 minutes and 3 at 20 minutes)

TABLE 2.12. *Sports Injuries Trial. Cross-classification of four dichotomized shivering measurements (at 5, 10, 15, and 20 minutes).*

(5 mins,	(15 mins, 20 mins)			
10 mins)	(0,0)	(0,1)	(1,0)	(1,1)
		Placebo arm		
(0,0)	37	11	8	16
(0,1)	6	2	6	23
(1,0)	1	0	0	1
(1,1)	2	0	4	21
		Treatment arm		
(0,0)	59	10	4	9
(0,1)	10	1	11	22
(1,0)	0	0	0	0
(1,1)	1	0	2	10

is a sampling zero because it is possible for a patient to go from being completely asleep to awake with eyes open, but rather unlikely. On the other hand, the zeros in the lower left hand corner of the table are all structural zeros because once a patient reached a certain level of consciousness, he could never return to a lower level. The longitudinal nature of the data is seen in Table 2.12, where the cross-classification of four dichotomized shivering measures, at 5, 10, 15, and 20 minutes, is shown.

2.9 Age Related Macular Degeneration Trial

These data arise from a randomized multi-centric clinical trial comparing an experimental treatment (interferon-α) to a corresponding placebo in the treatment of patients with age-related macular degeneration. Throughout the analyses done, we focus on the comparison between placebo and the highest dose (6 million units daily) of interferon-α (Z), but the full results of this trial have been reported elsewhere (Pharmacological Therapy for Macular Degeneration Study Group 1997). Patients with macular degeneration progressively lose vision. In the trial, the patients' visual acuity was assessed at different time points (4 weeks, 12 weeks, 24 weeks, and 52 weeks) through their ability to read lines of letters on standardized vision charts. These charts display lines of 5 letters of decreasing size, which the patient must read from top (largest letters) to bottom (smallest letters). Each line with at least 4 letters correctly read is called one 'line of vision.'

TABLE 2.13. *Age Related Macular Degeneration Trial. Loss of at least 3 lines of vision at 1 year according to loss of at least 2 lines of vision at 6 months and according to randomized treatment group (placebo* versus *interferon-α).*

	12 months			
	Placebo		Active	
6 months	0	1	0	1
No event (0)	56	9	31	9
Event (1)	8	30	9	38

TABLE 2.14. *Age Related Macular Degeneration Trial. Mean (standard error) of visual acuity at baseline, at 6 months and at 1 year according to randomized treatment group (placebo* versus *interferon-α).*

Time point	Placebo	Active	Total
Baseline	55.3 (1.4)	54.6 (1.3)	55.0 (1.0)
6 months	49.3 (1.8)	45.5 (1.8)	47.5 (1.3)
1 year	44.4 (1.8)	39.1 (1.9)	42.0 (1.3)

The patient's visual acuity is the total number of letters correctly read. The primary endpoint of the trial was the loss of at least 3 lines of vision at 1 year, compared to their baseline performance (a binary endpoint). The secondary endpoint of the trial was the visual acuity at 1 year (treated as a continuous endpoint). Buyse and Molenberghs (1998) examined whether the patient's performance at 6 months could be used as a surrogate for their performance at 1 year with respect to the effect of interferon-α. They looked at whether the loss of 2 lines of vision at 6 months could be used as a surrogate for the loss of at least 3 lines of vision at 1 year (Table 2.13). They also looked at whether visual acuity at 6 months could be used as a surrogate for visual acuity at 1 year.

Table 2.14 shows the visual acuity (mean and standard error) by treatment group at baseline, at 6 months, and at 1 year. Visual acuity can be measured in several ways. First, one can record the number of letters read. Alternatively, dichotomized versions (at least 3 lines of vision lost, or at least 3 lines of vision lost) can be used as well. Therefore, these data will be useful to illustrate methods for the joint modeling of continuous and binary outcomes, with or without taking the longitudinal nature into account. In addition, though there are 190 subjects with both month 6 and month 12 measurements available, the total number of longitudinal profiles is 240, but only for 188 of these have the four follow-up measurements been made.