No. 858

Evaluating Treatments of Rapidly Progressive Glomerulonephritis when the Response is Potentially Non-ignorably Missing

by

Ryoko Tobiishi , Yuichiro Kanazawa ,
Hideto Takahashi , Naoto Yamaguchi ,
Kunihiro Yamagata , and Tetsuo Koyama

April 2000

Evaluating Treatments of Rapidly Progressive Glomerulonephritis when the Response is Potentially Non-ignorably Missing

Ryoko Tobiishi[*1], Yuichiro Kanazawa[*2], Hideto Takahashi[*3], Naoto Yamaguchi[*4], Kunihiro Yamagata[*4], and Tetsuo Koyama[*4]

Doctoral Program in Policy and Planning Sciences, University of Tsukuba[*1]

Institute of Policy and Planning Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573, Japan[*2]

Institute of Community Medicine, University of Tsukuba[*3]

Institute of Clinical Medicine, University of Tsukuba[*4]

Abstract

Rapidly progressive glomerulonephritis (RPGN) is a life-threatening nephritis with a rapid decline in renal function. In this paper, we examine the three alternative treatments relative to the most commonly-used oral steroid for the two prevalent forms of RPGN, called PautiCrGN and MPA, in Japan. The response "Exitus" is potentially non-ignorably missing because the data were reported by those who administered the treatment. We use the method proposed in Ibrahim (1996) for estimating parameters in binomial logistic models when the missing response is potentially nonignorable. For PautiCrGN patients, we found that the three alternative treatments did not seem to affect the survival of the patients. MPA patients who underwent the steroid-pulse or the cyclophosphamide with the oral steroid, seemed less likely to have survived compared with those who received the oral steroid only. For PautiCrGN patients, their exitus was not non-ignorably missing at first sight, and its missingness increased when the patients

received the combined treatment. If, however, steroidpulse and/or cyclophosphamide were additionally prescribed as the patients' conditions deteriorated, then the observed correlation above may prove the existence of non-ignorably missing exitus instead, because the exitus of serious patients might have been under-reported precisely because of their conditions.

Key words: EM algorithm, PautiCrGN, MPA, non-ignorable, logistic model

1 Introduction

Rapidly progressive glomerulonephritis (RPGN) is a life-threatening nephritis with a rapid decline in renal function. According to the Research Group of Progressive Renal Lesion (RGPRL) funded by the Ministry of Health and Welfare, the number of patients with RPGN during April 1994 through March 1995 increased by 60 percent from the comparable period ending March 1991. Discovery of the anti-neutrophil cytoplasmic antibody (ANCA), which enables its early diagnosis, combined with progresses in several methods of treatment have made it possible for some RPGN patients to recover their renal functions at least partially, however.

There are several treatments currently available with possible therapeutic effect: the oral steroid; the steroidpulse; the immunosuppressant with the cyclophosphamide; the plasma exchange; the combined treatments based mainly on the steroid. In this paper, we wish to find out effective method(s) of treatment relative to the most commonly-used oral steroid based on the most comprehensive data on RPGN currently available in Japan.

There are two definitions of RPGN. One is the clinical: Rapidly progress-

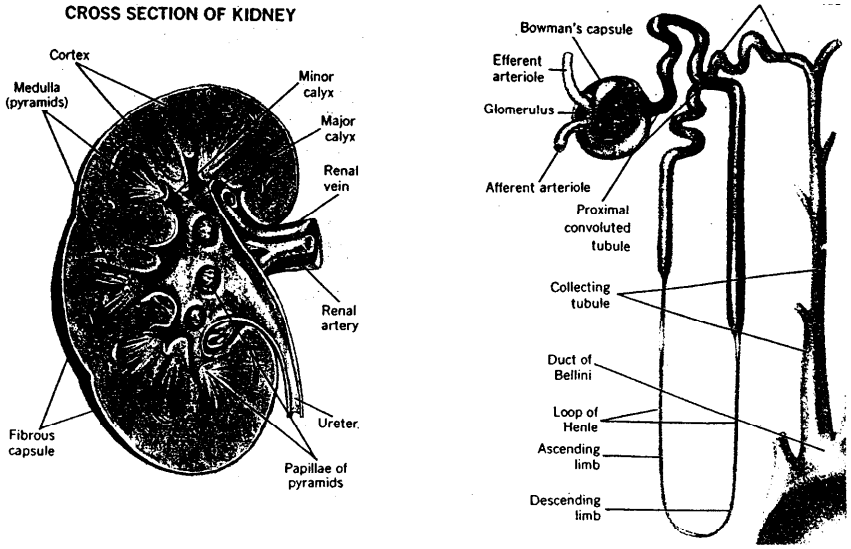

Figure 1: Kidney: The kidney is a pair of organs that eliminate wastes from the body as urine. The tissue portion of the kidney consists of two principal layers, an outer layer called the cortex and an inner layer called the medulla. The glomerulus is the capillary tuft which is located in the cortex, and the Bowman's capsule is a saclike double membrane which encloses the glomerulus.

gressing renal failure in a period of several weeks to months accompanied with nephritic urine such as hematuria (mainly microscopic but sometimes macroscopic), proteinuria, erythrocyte cost and granulated cylinder. The other is the pathological: Crescentic glomeruli are being formed in greater than 50 percent of circumferential involvement of the Bowman's capsule, see figure 2; further more than 50 percent of glomeruli have crescentic shape. The crescentic glomerulus may be defined as an aggregation of cells which can occupy a small or large segment of Bowman's space, and can extend virtually to obliterate the glomerulus.

According to Glassock's definition (1996), RPGN is classified into four major categories: the primary glomerular diseases; those associated with multisystem disease; those associated with infectious disease; those associated with medications. We are hoping to show the most effective treatment for

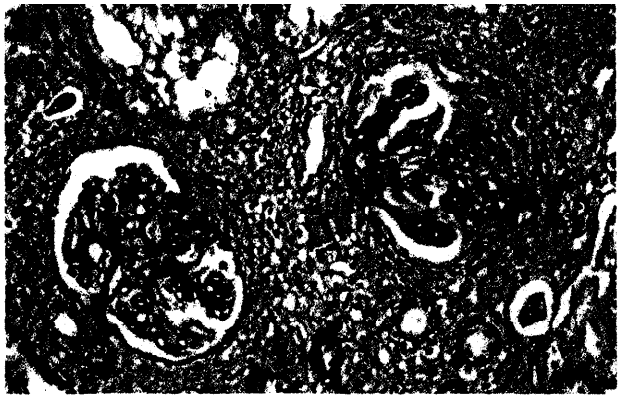

Figure 2: Crescentic glomeruli in the Bowman's capsule

the pauti-immune crescentic glomerulonephritis with ANCA (PautiCrGN) of the primary glomerular diseases and microscopic polyangiitis (MPA) of those associated with multisystem disease, two of the most prevalent forms of RPGN with the numbers of patients in our data being 237 and 127 respectively, far exceeding the third prevalent systemic lupus erythematosus with 42 patients. PautiCrGN and MPA are the nephritis thought to be caused by the existence of ANCA in the system, while ANCA itself is the antibody against the neutrophil cytoplasm. For the two diseases, the steroid and the cyclophosphamide are used to control the production of ANCA or activity of the neutrophil cytoplasm.

This paper is organized as follows. In section 2, the detailed explanation of the data is given. The method is described in section 3 and further details in appendix A. In section 4, the result is given, and in section 5 we examine the result.

2 Data

The aforementioned RGPRL conducted the survey on all the patients with RPGN in Japan retrospectively in August 1997. Following the Japanese Society of Nephrology guidelines, only the clinical definition is used to screen

the patients in this study. Those who were positively identified were all included.

The mail-in survey was sent to each one of the 365 renal special hospitals throughout the country. It asked them to fill-out the specified form based on their hospital records from January 1989 through August 1997. Of those hospitals, 134 responded. In 1998, the same forms were sent to those hospitals whose answers were incomplete in the first survey. If the responses in the second survey were somehow different, we would use responses in the second survey, assuming that the responses in the second survey were presumably more carefully filled out. The total number of patients was 715, and no patients were counted twice.

Treatments

In this data, patients with PautiCrGN and MPA respectively received 98 and 63 distinctive treatments altogether according to the progression of RPGN, age, and/or gender. However, most of these treatments included one of the four widely applied treatments as shown in table 1: 219 patients with PautiCrGN and 112 patients with MPA underwent at least one of the four treatments. The treatment 1 indicates the oral steroid, treatment 2 the oral steroid and steroidpulse, treatment 3 the oral steroid and cyclophosphamide, and treatment 4 the oral steroid, steroidpulse and cyclophosphamide as shown in table 1.

We will principally evaluate the relative effectiveness of these four treatments. There were eleven and nine confirmed patients that did not receive any of the four treatments for patients with PautiCrGN and MPA respectively. We excluded them from our analysis. Patients with missing treatment covariates were left included in our analysis because the likelihood of their

Table 1: Treatments

Treatment	1	2	3	4	5	other Tr	missing	total
oral steroid	1	1	1	1	0	0		
steroidpulse	0	1	0	1	0	0		
cyclophosphamide	0	0	1	1	1	0		
PautiCrGN	28	107	24	60	1	10	7	237
MPA	12	36	19	45	0	9	6	127

0; no, 1; yes

other Tr stands for other treatment.

receiving one of the four treatments was very high. Therefore we analyze 226 and 118 patients with PautiCrGN and MPA respectively.

To see the correlations between the "Treatments" and four other covariates—age, the serum creatinine level, the existence of the lesion of lung, and gender—that could potentially affect both on the response and the treatment, we run logistic regression using only patients whose data were completely recorded.

Since the treatments 1, 2, 3, and 4 are devoid of structure and should not be ordered, we fitted the multinomial logistic model setting the treatment 1 as a baseline and using two dummy variables each for age and serum creatinine level to account for possible non-linear effect of these variable in the choice of treatment. The "Age2" indicates 1 for 60 $\leq$age$<$ 69, 0 for others, and "Age3" 1 for 70 $\leq$ age, 0 for others. The "Cr2" indicates 1 for 3 $\leq$Cr$<$ 6, 0 for others, and "Cr3" 1 for 6 $\leq$Cr, 0 for others. Tables 2 and 3 illustrate the elimination process.

We found "Age3" for patients with PautiCrGN and "Cr3" for patients

Table 2: The estimation for patients with PautiCrGN

	Response	Constant	Age2	Age3	Cr2	Cr3	Lesion	Gender
Step1	$\frac{Treatment2}{Treatment1}$	1.844	-0.539	-0.662	-0.109	0.393	0.312	-0.125
		(1.864)	(-0.873)	(-1.054)	(-0.195)	(0.680)	(0.652)	(-0.273)
	$\frac{Treatment3}{Treatment1}$	-0.509	-0.716	-0.906	0.466	-0.379	0.323	0.468
		(-0.386)	(-0.953)	(-1.141)	(0.670)	(-0.455)	(0.521)	(0.753)
	$\frac{Treatment4}{Treatment1}$	0.831	-0.270	-1.364	0.409	0.251	0.679	-0.0425
		(0.769)	(-0.418)	(-1.929)	(0.672)	(0.381)	(1.304)	(-0.085)
Step2	$\frac{Treatment2}{Treatment1}$	1.562	-0.584	-0.851	-0.096	0.367	0.559	
		(2.778)	(-0.970)	(-1.402)	(-0.180)	(0.667)	(1.199)	
	$\frac{Treatment3}{Treatment1}$	0.418	-0.752	-1.208	0.255	-0.594	0.468	
		(0.622)	(-1.036)	(-1.567)	(0.385)	(-0.741)	(0.768)	
	$\frac{Treatment4}{Treatment1}$	0.863	-0.490	-1.597	0.369	0.254	0.838	
		(1.425)	(-0.782)	(-2.358)	(0.631)	(0.406)	(1.654)	
Step3	$\frac{Treatment2}{Treatment1}$	1.517	-0.592	-0.852		0.421	0.553	
		(3.015)	(-0.988)	(-1.403)		(0.895)	(1.193)	
	$\frac{Treatment3}{Treatment1}$	0.545	-0.724	-1.215		-0.744	0.494	
		(0.932)	(-1.003)	(-1.577)		(-1.057)	(0.816)	
	$\frac{Treatment4}{Treatment1}$	1.056	-0.453	-1.597		0.022	0.870	
		(2.001)	(-0.726)	(-2.361)		(0.042)	(1.726)	
Step4	$\frac{Treatment2}{Treatment1}$	1.590	-0.533	-0.719			0.599	
		(3.194)	(-0.897)	(-1.208)			(1.312)	
	$\frac{Treatment3}{Treatment1}$	0.485	-0.815	-1.355			0.397	
		(0.836)	(-1.135)	(-1.786)			(0.666)	
	$\frac{Treatment4}{Treatment1}$	1.189	-0.558	-1.679			0.811	
		(2.3)	(-0.907)	(-2.540)			(1.638)	
Step5	$\frac{Treatment2}{Treatment1}$	1.261		-0.381			0.578	
		(4.101)		(-0.867)			(1.268)	
	$\frac{Treatment3}{Treatment1}$	0.010		-0.864			0.360	
		(0.027)		(-1.407)			(0.606)	
	$\frac{Treatment4}{Treatment1}$	0.847		-1.326			0.788	
		(2.581)		(-2.566)			(1.595)	
Step6	$\frac{Treatment2}{Treatment1}$	1.447		-0.294				
		(5.207)		(-0.681)				
	$\frac{Treatment3}{Treatment1}$	0.118		-0.811				
		(0.343)		(-1.337)				
	$\frac{Treatment4}{Treatment1}$	1.119		-1.206				
		(3.887)		(-2.379)				

Asymptotic t-values in parentheses.

Table 3: The estimation for patients with MPA

	Response	Constant	Age2	Age3	Cr2	Cr3	Lesion	Gender
Step1	$\frac{Treatment2}{Treatment1}$	-0.045	-0.785	-0.979	1.996	1.620	-0.116	0.548
		(-0.027)	(-0.635)	(-0.772)	(1.591)	(1.864)	(-0.135)	(0.721)
	$\frac{Treatment3}{Treatment1}$	0.656	-0.517	0.0449	0.906	-0.972	-1.555	0.578
		(0.363)	(-1.088)	(-1.058)	(1.576)	(0.913)	(0.462)	(0.402)
	$\frac{Treatment4}{Treatment1}$	1.091	-1.302	-1.295	1.872	0.737	0.385	0.295
		(0.699)	(-1.088)	(-1.058)	(1.576)	(0.913)	(0.462)	(0.402)
Step2	$\frac{Treatment2}{Treatment1}$	1.153	-1.356	-1.064	1.222	1.679	-0.148	
		(0.950)	(-1.139)	(-0.855)	(1.223)	(1.997)	(-0.187)	
	$\frac{Treatment3}{Treatment1}$	1.478	-0.903	0.339	0.320	-0.506	-1.452	
		(1.166)	(-0.685)	(0.252)	(0.313)	(-0.513)	(-1.714)	
	$\frac{Treatment4}{Treatment1}$	1.574	-1.717	-1.367	1.366	0.879	0.473	
		(1.343)	(-1.480)	(-1.128)	(1.467)	(1.099)	(0.608)	
Step3	$\frac{Treatment2}{Treatment1}$	0.469	-0.605		1.301	1.669	-0.260	
		(0.575)	(-0.861)		(1.311)	(1.983)	(-0.332)	
	$\frac{Treatment3}{Treatment1}$	1.662	-1.081		0.219	-0.443	-1.346	
		(2.161)	(-1.363)		(0.216)	(-0.451)	(-1.612)	
	$\frac{Treatment4}{Treatment1}$	0.761	-0.790		1.470	0.872	0.291	
		(0.977)	(-1.163)		(1.594)	(1.094)	(0.383)	
Step4	$\frac{Treatment2}{Treatment1}$	0.336	-0.605		1.277	1.581		
		(0.475)	(-0.862)		(1.291)	(1.978)		
	$\frac{Treatment3}{Treatment1}$	1.148	-1.095		0.121	-0.944		
		(1.739)	(-1.409)		(0.122)	(-1.018)		
	$\frac{Treatment4}{Treatment1}$	0.927	-0.780		1.490	0.957		
		(1.425)	(-1.149)		(1.620)	(1.255)		
Step5	$\frac{Treatment2}{Treatment1}$	5.1792e-16			1.386	1.609		
		(8.971e-16)			(1.416)	(2.022)		
	$\frac{Treatment3}{Treatment1}$	0.606			0.310	-0.894		
		(1.194)			(0.317)	(-0.975)		
	$\frac{Treatment4}{Treatment1}$	0.511			1.629	0.993		
		(0.989)			(1.793)	(1.313)		
Step6	$\frac{Treatment2}{Treatment1}$	0.560				1.050		
		(1.263)				(1.490)		
	$\frac{Treatment3}{Treatment1}$	0.693				-0.981		
		(1.601)				(-1.117)		
	$\frac{Treatment4}{Treatment1}$	1.216				0.288		
		(3.022)				(0.421)		

Asymptotic t-values in parentheses.

with MPA in step5 of table 3 were statistically significant with asymptotic t-values exceeding 2 in absolute terms. From the estimated regression coefficients, the PautiCrGN patients under seventy years are more likely to have undergone the treatment 4 compared with treatment 1. For patients with MPA, patients with creatinine level ≥ 6 are more likely to have undergone the treatment 2 relative to treatment 1. Similarly, we fitted the multinomial logistic model using the original undiscretized age and serum creatinine, and we found the result remained basically the same. Therefore we are forced to believe that these patients were non-randomly assigned to these treatments.

Exitus

In our analysis, the binary exitus is a response variable indicating whether the patient is survived(= 1) or dead(= 0) as in table 4 and figures 3 and 4.

Table 4: Exitus

	1	0	missing	total
PautiCrGN	162	56	8	226
MPA	67	48	3	118

Age, Serum Creatinine, Lesion of lung, and Gender

As we already mentioned, we included four covariates other than the aforementioned four treatments. They were age, the serum creatinine level, the existence of the lesion of lung and gender. Clinical features of the 226 patients with PautiCrGN and 118 patients with MPA were summarized in table 5, and figures 5 to 12.

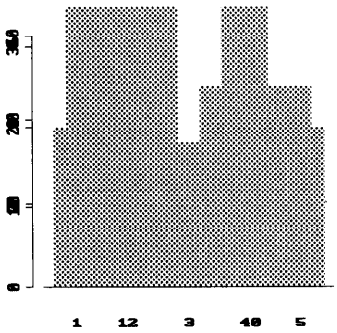

Figure 3: PautiCrGN

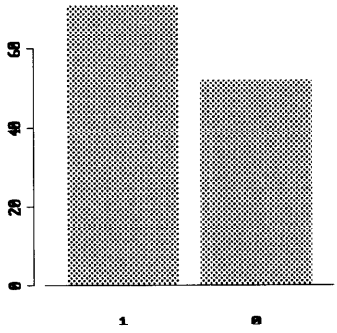

Figure 4: MPA

Table 5: Clinical features

Age	10-19	20-29	30-39	40-49	50-59	60-69
PautiCrGN	9	7	6	10	34	91
MPA	1	1	1	4	20	48
	70-79	80-89	missing	total		
	55	13	1	226		
	38	5	0	118		
Serum Creatinine	Cr< 3	3 ≤Cr< 6	6 ≤Cr	missing	total	
PautiCrGN	64	83	76	3	226	
MPA	33	35	48	2	118	
Lesion of lung	Yes	No	missing	total		
PautiCrGN	95	130	1	226		
MPA	75	43	0	118		
Gender	male	female	missing	total		
PautiCrGN	93	122	11	226		
MPA	50	61	7	118		

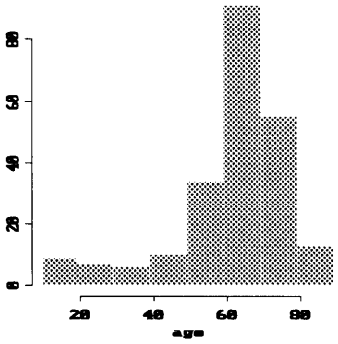

Figure 5: Age: PautiCrGN

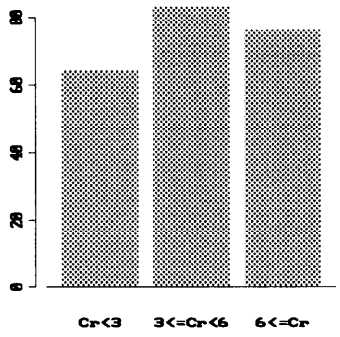

Figure 6: Creatinine: PautiCrGN

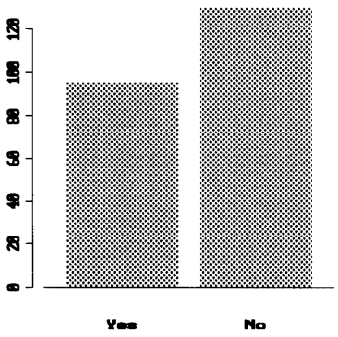

Figure 7: Lesion of lung: PautiCrGN

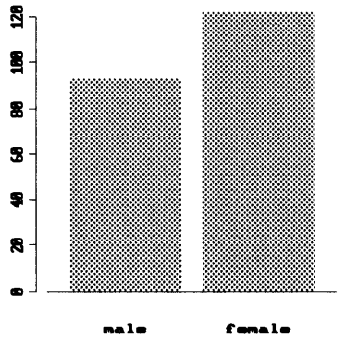

Figure 8: Gender: PautiCrGN

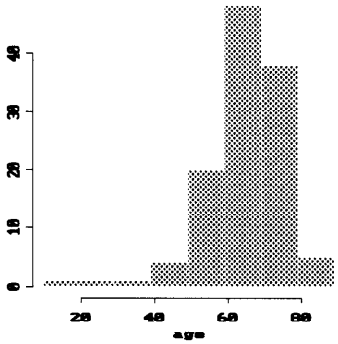

Figure 9: Age: MPA

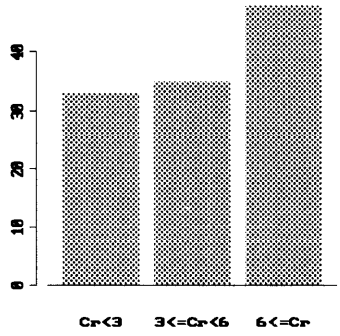

Figure 10: Creatinine: MPA

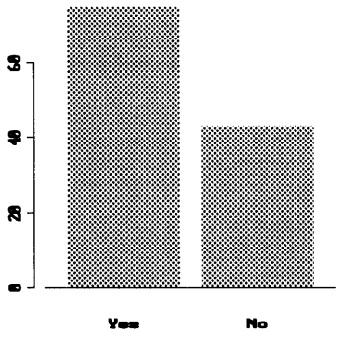

Figure 11: Lesion of lung: MPA

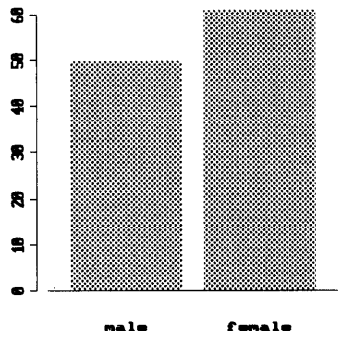

Figure 12: Gender: MPA

Missing-value Problem

Tables 6 and 7 illustrate the missing-value problem in the 226 PautiCrGN and 118 MPA patients. There were missing-values in the response and all the covariates as seen in table 6. For the two prevalent forms of RPGN, the figures in table 7 show the percentage of patients without a single missing-value in both of the response and covariates. Ignoring the remaining 10.6% and 15.3% of the data could conceivably cause serious bias in the analysis.

Table 6: The magnitude of missing-value problem for the patients with PautiCrGN and with MPA

	Exitus	Treatment	Age	Cr	Lesion	Gender
PautiCrGN	3.5%	3.1%	0.4%	1.32%	0.4%	4.86%
MPA	2.5%	5.1%	0%	1.69%	0%	5.9%

There were eight and three missing exitus or 3.5% and 2.5% for patients with PautiCrGN and with MPA respectively.

Compared with missing value problem in the covariates, we think missing

Table 7: The ratio of patients without a single missing value

PautiCrGN	MPA
89.4%	84.7%

response could be more harmful. When non-response is related to the missing values of the variables, the non-response is called non-ignorable. In this data, non-ignorable missing response could have happened, for example, when the treatment did not produce the desired results and patients died. When there is non-ignorable non-response, ignoring the missing data mechanism could cause serious bias.

3 Method

In this section, we summarize the method for estimating parameters in binomial logistic regression models when the response variable $\mathbf{Y}$ is missing and the missing data mechanism is potentially nonignorable, and the covariates are fully observed. When both the response and the covariates are missing, we impute the missing covariate with the value of the similar patterned patient.

The complete-data model consists of the joint distribution of the response variable $\mathbf{Y}$ and the missing data indicator $\mathbf{R}$. Since the covariates $\mathbf{X}$ are fully observed, they are treated as fixed throughout. We express the joint distribution $\mathbf{R}$ and $\mathbf{Y}$ by specifying the conditional distributions $(\mathbf{Y}|\boldsymbol{\beta})$ and $(\mathbf{R}|\mathbf{Y}, \boldsymbol{\alpha})$, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are assumed to be distinct sets of indexing parameters for their respective distributions.

Suppose $y_1, \ldots, y_n$ are independent observations, where each y_i has a

Bernoulli distribution with success probability π_i, $i = 1, \ldots, n$. Further, let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$ denote the $1 \times p$ observed vector of covariates for the ith observation, $\mathbf{X}$ is an $n \times p$ matrix of cavariates, and let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ denote the corresponding $p \times 1$ column vector of regression coefficients. A $x_{i1} = 1$ is included in $\boldsymbol{x}_i$. We use a logistic regression model for the y_i's with the likelihood for $y_i|\boldsymbol{x}$ is given by

$$\begin{aligned} L_{y_i}(\boldsymbol{\beta}) &= \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i}(1-\pi_i) \\ &= \exp\left[y_i \boldsymbol{x}_i\boldsymbol{\beta} - \log\left\{1 + \exp(\boldsymbol{x}_i\boldsymbol{\beta})\right\}\right], \end{aligned} \tag{3.1}$$

where the logistic link function $\log\{\pi_i/(1-\pi_i)\} = \boldsymbol{x}_i\boldsymbol{\beta}$ is used. One important property of the logistic function not shared by the other link functions is that differences on the logistic scale can be estimated regardless of whether the data are sampled prospectively or retrospectively. The missing data indicator for the ith response y_i can be written as

$$r_i = \begin{cases} 1 & \text{if } y_i \text{ is observed,} \\ 0 & \text{if } y_i \text{ is missing,} \end{cases}$$

for $i = 1, \ldots, n$. The vector $\boldsymbol{r} = (r_1, \ldots, r_n)^T$ is $n \times 1$ column vector of missing data indicators. We specify a logistic regression model for the r_i's. Let $\boldsymbol{z}_i = (\boldsymbol{x}_i, y_i)$ and let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{p+1})^T$ be a $(p+1) \times 1$ column vector of indexing parameters for r_i. The likelihood for r_i is

$$\begin{aligned} L_{r_i|y_i}(\boldsymbol{\alpha}) &= \left(\frac{p_i}{1-p_i}\right)^{r_i}(1-p_i) \\ &= \exp\left[r_i \boldsymbol{z}_i\boldsymbol{\alpha} - \log\{1 + \exp(\boldsymbol{z}_i\boldsymbol{\alpha})\}\right], \end{aligned} \tag{3.2}$$

where $p_i = \Pr\{r_i = 1|\boldsymbol{z}_i, \boldsymbol{\alpha}\}$ and $\log\{p_i/(1-p_i)\} = \boldsymbol{z}_i\boldsymbol{\alpha}$. We see that if $\alpha_{p+1} \neq 0$, then the missing data mechanism depends on y_i and is nonignorable. If $\alpha_{p+1} = 0$, then $f(r_i|\boldsymbol{z}_i, \boldsymbol{\alpha})$ does not depend on y_i, which implies that the missing data are missing at random and the missing data mechanism is

ignorable. If $\alpha_2 = \cdots = \alpha_{p+1} = 0$, then this implies that the missing data are missing completely at random.

Under the assumption that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are distinct sets of indexing parameters, the complete-data log-likelihood for all of the observations can be decomposed from (3.1) and (3.2) as

$$\begin{aligned} l(\boldsymbol{\gamma}) &= \sum_{i=1}^{n} l(\boldsymbol{\gamma}; \boldsymbol{x}_i, y_i, r_i) = \sum_{i=1}^{n} \{l_{y_i}(\boldsymbol{\beta}) + l_{r_i|y_i}(\boldsymbol{\alpha})\} \\ &= \sum_{i=1}^{n} [y_i \boldsymbol{x}_i \boldsymbol{\beta} - \log\{1 + \exp(\boldsymbol{x}_i \boldsymbol{\beta})\} \\ &\qquad + r_i \boldsymbol{z}_i \boldsymbol{\alpha} - \log\{1 + \exp(\boldsymbol{z}_i \boldsymbol{\alpha})\}], \end{aligned} \tag{3.3}$$

where $\boldsymbol{\gamma} = (\beta_1, \ldots, \beta_p, \alpha_1, \ldots, \alpha_{p+1})^T$ is a $(2p+1) \times 1$ column vector of logistic regression parameters and $l(\boldsymbol{\gamma}; \boldsymbol{x}_i, y_i, r_i)$ is the contribution to the complete data log-likelihood of the ith observation. The complete data model in (3.3) essentially treats the y_i's as missing covariates in the model for $(r_i|\boldsymbol{z}_i, \boldsymbol{\alpha})$. Thus following Ibrahim (1990), the maximum likelihood estimates of $\boldsymbol{\gamma}$ can be obtained via the EM algorithm by maximizing the expected log-likelihood. A detailed explanation of EM algorithm is in appendix A.

4 Result

With patients whose data were completely recorded

We first analyze the exitus as response using only patients whose data were completely recorded. This serves two purposes: This preliminary analysis hopefully give us an insight for eliminating some of the redundant covariates; The algorithm to be introduced in appendix A sometimes lead to inestimable parameters as Ibrahim and Lipsitz (1996, 1075) stated: "It is not clear how to characterize the set of all estimable parameters for this class of models". Our

experiences suggest that this problem is more pronounced when regressing $\mathbf{R}$ on categorical $\mathbf{X}$ and $\mathbf{Y}$ with many columns in $\mathbf{X}$. Therefore it is crucial to keep only the necessary covariates. We used two dummy variables each for age and serum creatinine as described in section 2. Tables 8 and 9 illustrate the elimination process for the PautiCrGN and MPA patients respectively.

PautiCrGN

From the t-values of estimated coefficients, we chose model 8 in table 8 as our final model. In the model, we found "Age3" and "Lesion of lung" were significant, but the treatments did not seem to be correlated with survival of the PautiCrGN patients. From the estimated regression coefficients, the odds that the patients under seventy who underwent "Treatment 1" (oral steroid) survived was about $\exp(1.045) = 2.843$ times the odds of those over seventy. The odds that the patients without "Lesion of lung" survived was about $\exp(1.297) = 3.658$ times the odds of those with.

MPA

Patients with MPA who underwent "Treatment 2" or "Treatment 4" were less likely to have survived, compared with "Treatment 1" as seen in table 9. Since we found a correlation between one of the treatments and the creatinine level "Cr3" in section 2, we would prefer to leave "Cr3" in the model to control the effect of treatments on exitus. So we chose the model 4 in table 9 as our final model. The odds that the patients who underwent "Treatment 2" survived was about $\exp(-1.641) = 0.194$ times the odds of those patients with "Treatment 1". Also the odds that patients who underwent "Treatment 4" survived was about $\exp(-1.210) = 0.298$ times the odds of those patients with "Treatment 1".

With all the patients

Next we try to find out which treatments were effective if the variable "Exitus" was possibly non-ignorably missing. To determine a suitable model for the missing data mechanism, we followed suggestion of Ibrahim and Lipsitz (1996, 1075): "One can start with a "full" model, and then do variable selection on $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ and keeping the covariates for the model $\mathbf{Y}|\mathbf{X}$ fixed. One can then use the likelihood ratio or AIC criterion to evaluate the fit of each model".

PautiCrGN

From the analysis of the PautiCrGN patients whose data were completely recorded, we chose to include the variables "Age3" and "Lesion of lung" when regressing $\mathbf{R}$. Further, we chose to employ the treatments and exitus as covariates for they might be related to missingness in "Exitus" in the following manner: The treatment did not produce the desired results and the patient died and her/his exitus was not reported because s/he died.

For the patients with PautiCrGN, "full" model with the missing data indicator $\mathbf{R}$ being regressed on the six covariates—"Treatment 2", "Treatment 3", "Treatment 4", "Age3", "Lesion of lung", "Exitus"—did not produce estimable parameters. The model closest to the "full" model with two of the three treatments did not either. Therefore we list the third-tier models with the $\mathbf{R}$ being regressed on the four covariates—one of "Treatments 2, 3, and 4", "Age3", "Lesion of lung", "Exitus"—and the models whose covariates were their subsets. Results of $\mathbf{Y}|\mathbf{X}$ is in table 10 and those of $\mathbf{R}|\mathbf{Z} = \mathbf{R}|(\mathbf{X}, \mathbf{Y})$ is in table 11.

Based on the AIC and the Deviance, we selected model 12 in table 10 as our final model. In the model, we found "Age3" and "Lesion of lung" were

significant, "Age2" and "Cr3" were marginally significant with the associated absolute t-values between one and two, but the treatments did not seem to be correlated with survival of the patients. From the estimated regression coefficients, the younger PautiCrGN patients without the lesion of lung and with low creatinine level were more likely to have survived. The odds that the patients under seventy who underwent "Treatment 1" (oral steroid) survived was about $\exp(1.8365) = 6.2745$ times the odds of those over seventy. The odds that the patients without "Lesion of lung" survived was about $\exp(1.0263) = 2.7907$ times the odds of those with "Lesion of lung".

Missingness in "Exitus" was likely to have increased when the patients received "Treatment 4". Therefore this non-response can be characterized as ignorable. We will discuss this issue in section 5. The odds that patients who underwent "Treatment 4" had their exitus missing was about $\exp(1.4231) = 4.15$ times the odds of the those who underwent "Treatment 1".

As we can see in appendix B, when the response is ignorably missing, the resulting logistic regression $\mathbf{Y}|\mathbf{X}$ parameter estimate $\hat{\boldsymbol{\beta}}$ in (3.1) and its associated standard error are equivalent to those obtained from completely observed data. This result suggests that, when missing value mechanism were presumed but turned out not to be non-ignorable, then the result of $\mathbf{Y}|\mathbf{X}$ in the simultaneous logistic regression of $\mathbf{Y}|\mathbf{X}$ and $\mathbf{R}|\mathbf{Z} = \mathbf{R}|(\mathbf{X}, \mathbf{Y})$ should be close to the result of $\mathbf{Y}|\mathbf{X}$ when $\mathbf{Y}$ are completely observed. Since model 12 in table 10 corresponds to $\mathbf{Y}|\mathbf{X}$ in the simultaneous logistic regression and model 1 in table 8 corresponds to $\mathbf{Y}|\mathbf{X}$ when $\mathbf{Y}$ are completely observed, their coefficients and t-values should be nearly the same. This is exactly what we observed. The same can not be said when missing values are in covariates.

MPA

For patients with MPA, we included the variables "Treatments 2 and 4" partly because they were statistically significant in the analysis of those completely recorded MPA patients and partly because they might be related to missingness in "Exitus" for the same reason as for the PautiCrGN patients.

However we could not obtain regression coefficients when the missing data mechanism was non-ignorable. This was because all the models in regressing $\mathbf{R}$, the weights corresponding to missing values in (A.1) were approximately 0 for $y_i = 0$ in the E-step of EM algorithm. Since the category with $\mathbf{R} = 0$ and $\mathbf{Y} = 0$ had no count as shown in figure 13, there did not exist a uniquely maximum likelihood estimate. For this case, we had no choice but to resort to the analysis with patients whose data were completely recorded.

Figure 13: two-way contingency table when regressing $\mathbf{R}$ on only $\mathbf{Y}$

		R		
		1	0	
Y	1	67	3	70
	0	48	0	48
		115	3	118

5 Discussion

Treatment on Exitus

In our analysis, the PautiCrGN patients did not show improvement with "Treatments 2, 3, and 4" relative to "Treatment 1". We also found that MPA patients who underwent "Treatment 2" or "Treatment 4" were less likely to

Table 8: The regression **Y** on **X** for PautiCrGN patients whose data were completely recorded

Model	Constant	Treatment2	Treatment3	Treatment4	Age2	Age3	Cr2	Cr3	Lesio
1	3.743	-0.6167	-0.9207	-0.9133	-1.0175	-1.8516	-0.1795	-0.9211	-1.04
	(2.9284)	(-0.7332)	(-0.8738)	(-1.0097)	(-1.4301)	(-2.627)	(-0.2785)	(-1.5359)	(-2.23
2	3.6453	-0.6143	-0.9267	-0.9238	-1.0358	-1.85		-0.8151	-1.05
	(2.970)	(-0.7313)	(-0.881)	(-1.0228)	(-1.4605)	(-2.6233)		(-1.7865)	(-2.25
3	3.139		-0.4083	-0.4018	-1.0106	-1.8289		-0.8163	-1.08
	(3.152)		(-0.5389)	(-0.7631)	(-1.4252)	(-2.5924)		(-1.7934)	(-2.34
4	3.115			-0.3417	-1.023	-1.825		-0.7927	-1.08
	(3.128)			(-0.6679)	(-1.445)	(-2.589)		(-1.755)	(-2.34
5	2.8999				-1.0832	-1.751		-0.6910	-1.20
	(3.021)				(-1.553)	(-2.549)		(-1.5556)	(-2.63
6	3.7324				-1.1092	-1.768		-0.7405	-1.24
	(5.521)				(-1.603)	(-2.589)		(-1.686)	(-2.76
7	3.006					-1.0192		-0.7685	-1.26
	(7.075)					(-2.35)		(-1.762)	(-2.80
8	2.686					-1.045			-1.29
	(7.140)					(-2.501)			(-3.03

Asymptotic t-values in parentheses.

Table 9: The regression **Y** on **X** for MPA patients whose data were completely recorded

Model	Constant	Treatment2	Treatment3	Treatment4	Age2	Age3	Cr2	Cr3	Lesion
1	4.094	-1.828	-0.887	-1.469	-1.064	-0.247	-1.271	-1.049	-0.899
	(2.771)	(-1.571)	(-0.655)	(-1.293)	(-1.504)	(-0.352)	(-1.696)	(-1.439)	(-1.424)
2	4.572	-2.266	-1.0486	-1.836	-1.244	-0.444	-1.258	-1.070	-0.858
	(3.395)	(-1.952)	(-0.784)	(-1.625)	(-1.805)	(-0.651)	(-1.731)	(-1.492)	(-1.378)
3	4.3066	-2.229	-1.0898	-1.803	-0.953		-1.226	-1.068	-0.933
	(3.373)	(-1.924)	(-0.815)	(-1.5976)	(-1.852)		(-1.695)	(-1.492)	(-1.528)
4	3.633	-1.641		-1.210	-0.917		-1.205	-1.0187	-0.887
	(4.182)	(-2.0885)		(-1.639)	(-1.803)		(-1.673)	(-1.447)	(-1.468)
5	3.313	-1.967		-1.416	-0.889		-0.539		-1.0855
	(4.083)	(-2.606)		(-1.951)	(-1.769)		(-1.012)		(-1.869)
6	3.146	-1.998		-1.478	-0.838				-1.082
	(4.049)	(-2.682)		(-2.053)	(-1.733)				(-1.869)
7	2.7699	-1.891		-1.422					-1.132
	(3.785)	(-2.5996)		(-2.0037)					(-1.978)
8	2.1203	-2.1203		-1.6503					
	(3.4701)	(-2.978)		(-2.378)					

Asymptotic t-values in parentheses.

Table 10: The regression **Y** on **X** for PautiCrGN patients in considering exitus possibly non-ignorably missing

Model	Constant	Treatment2	Treatment3	Treatment4	Age2	Age3	Cr2	Cr3	Lesi
1	3.6622	-0.61911	-0.97202	-0.96852	-0.99357	-1.8413	-0.1852	-0.8851	-1.05
	(2.8687)	(-0.73627)	(-0.92168)	(-1.0718)	(-1.3918)	(-2.6128)	(-0.28668)	(-1.4723)	(-2.25
2	3.6647	-0.61912	-0.96985	-0.96891	-0.99623	-1.8436	-0.18374	-0.88482	-1.05
	(2.8705)	(-0.73628)	(-0.91964)	(-1.0722)	(-1.3955)	(-2.6157)	(-0.28441)	(-1.4718)	(-2.2
3	3.7487	-0.61657	-0.91699	-0.90842	-1.0191	-1.8514	-0.17821	-0.92365	-1.04
	(2.9327)	(-0.73307)	(-0.87047)	(-1.0044)	(-1.4329)	(-2.6273)	(-0.27658)	(-1.5405)	(-2.2
4	3.4539	-0.62697	-1.092	-1.0554	-0.93582	-1.7679	-0.14975	-0.79248	-1.04
	(2.75)	(-0.74755)	(-1.0473)	(-1.1745)	(-1.3319)	(-2.5661)	(-0.23359)	(-1.3275)	(-2.25
5	3.7125	-0.61763	-0.94054	-0.9377	-1.0086	-1.8504	-0.18396	-0.90703	-1.04
	(2.9049)	(-0.73433)	(-0.89207)	(-1.0369)	(-1.4152)	(-2.6236)	(-0.28508)	(-1.5105)	(-2.2
6	3.7678	-0.61598	-0.90428	-0.88911	-1.0243	-1.8488	-0.17262	-0.9329	-1.03
	(2.9479)	(-0.73243)	(-0.85897)	(-0.98338)	(-1.4423)	(-2.6269)	(-0.26821)	(-1.5576)	(-2.2
7	3.7399	-0.61683	-0.92281	-0.9162	-1.0166	-1.8517	-0.1801	-0.91953	-1.04
	(2.9258)	(-0.73337)	(-0.87577)	(-1.013)	(-1.4285)	(-2.6267)	(-0.27939)	(-1.533)	(-2.23
8	3.7994	-0.61502	-0.8822	-0.84911	-1.0332	-1.8385	-0.15687	-0.94881	-1.02
	(2.974)	(-0.7315)	(-0.8392)	(-0.9407)	(-1.4588)	(-2.6228)	(-0.2445)	(-1.5878)	(-2.2(
9	3.7979	-0.61508	-0.8833	-0.8515	-1.0328	-1.8394	-0.15789	-0.94795	-1.02
	(2.9727)	(-0.73155)	(-0.8402)	(-0.94323)	(-1.458)	(-2.6232)	(-0.24603)	(-1.5861)	(-2.2(
10	3.7639	-0.6161	-0.90688	-0.89326	-1.0233	-1.8496	-0.17396	-0.93102	-1.03
	(2.9448)	(-0.73255)	(-0.86131)	(-0.98788)	(-1.4404)	(-2.6271)	(-0.2702)	(-1.5541)	(-2.2
11	3.7996	-0.61495	-0.8822	-0.84731	-1.0327	-1.8372	-0.15648	-0.94944	-1.02
	(2.9742)	(-0.73143)	(-0.8392)	(-0.9388)	(-1.4581)	(-2.6213)	(-0.24392)	(-1.589)	(-2.1
12	3.8032	-0.61489	-0.8795	-0.8432	-1.0341	-1.8365	-0.15432	- 0.9509	-1.02
	(2.9772)	(-0.7314)	(-0.8368)	(-0.9345)	(-1.4606)	(-2.6217)	(-0.2407)	(-1.5918)	(-2.1
13	3.7076	-0.6177	-0.9438	-0.9412	-1.0074	-1.85	-0.18421	-0.90476	-1.04
	(2.9012)	(-0.7345)	(-0.89504)	(-1.0409)	(-1.4131)	(-2.623)	(-0.28543)	(-1.5065)	(-2.2
14	3.804	-0.61483	-0.87905	-0.84105	-1.034	-1.8352	-0.15358	-0.95163	-1.02
	(2.9778)	(-0.73133)	(-0.8364)	(-0.93223)	(-1.4605)	(-2.6206)	(-0.23953)	(-1.5932)	(-2.1

Asymptotic t-values in parentheses.

Table 11: The regression **R** on **Y** and **X** for PautiCrGN patients in considering exitus possibly non-ignorably missing

Model	Constant	Treatment2	Treatment3	Treatment4	Age3	lesion	Exitus
1	2.2709	1.3716			0.50733	0.165	0.79506
	(0.8625)	(1.2078)			(0.37111)	(0.13982)	(0.2802)
AIC: 204.37, Deviance=174.37							
2	2.825		-0.68878		0.65724	0.15398	0.77521
	(0.9209)		(-0.59845)		(0.47685)	(0.11484)	(0.2372)
AIC: 205.86, Deviance=175.86							
3	4.1202			-1.3984	0.21029	0.14276	-0.083471
	(0.96857)			(-1.4624)	(0.13718)	(0.12668)	(-0.019543)
AIC: 204.36, Deviance=174.36							
4	1.7358				1.0303	0.5228	1.9557
	(0.82384)				(0.7254)	(0.38814)	(0.71708)
AIC: 201.63, Deviance=173.63							
5	2.691	1.3712			0.41606		0.3649
	(1.1601)	(1.2083)			(0.31552)		(0.13893)
AIC: 202.68, Deviance=178.68							
6	3.4378	1.3996				-0.02227	-0.40037
	(0.91268)	(1.2376)				(-0.02069)	(-0.1022)
AIC: 202.78, Deviance=174.78							
7	3.4804		-0.70383		0.50987		0.07183
	(1.2417)		(-0.61395)		(0.39436)		(0.02323)
AIC: 204.19, Deviance=176.19							
8	5.3339			-1.4184	0.02604		-1.3008
	(0.67006)			(-1.4732)	(0.01808)		(-0.16118)
AIC: 201.87, Deviance=173.87							
9	5.2622			-1.426		0.037752	-1.2307
	(0.77543)			(-1.526)		(0.035604)	(-0.17734)
AIC: 201.91, Deviance=173.91							
10	3.3655	1.4009					-0.32913
	(1.1544)	(1.2391)					(-0.10168)
AIC: 200.8, Deviance=174.8							
11	4.8617		-0.71836				-1.3213
	(0.70567)		(-0.62652)				(-0.18262)
AIC: 201.88, Deviance=175.88							
12	5.5235			-1.4231			-1.4913
	(0.75131)			(-1.5294)			(-0.19522)
AIC: 199.77, Deviance=173.77							
13	3.0779				0.61154		0.40677
	(1.3142)				(0.46214)		(0.15171)
AIC: 202.47, Deviance=176.47							
14	4.9585						-1.5359
	(0.5932)						(-0.17651)
AIC: 200.11, Deviance=176.11							

Asymptotic t-values in parentheses.

have survived, compared with those who received "Treatment 1". This lack of correlation in the former, and the existence of negative correlation in the latter between the exitus and some of the treatments were totally unexpected. We suspect that these results were obtained possibly because steroidpulse and/or cyclophosphamide were additionally prescribed for patients whose conditions did not improve with their initial oral steroid treatments.

The data used in our analysis were retrospectively collected from the hospital records and the investigators were not able to randomize the treatments. In an observational study such as this, there may be some confounding factors to create an apparent correlation between "Exitus" and the treatments. To establish cause and effect relationship between the treatments and the patients' prognoses—as this is what is usually required, a new study must be designed and carried out in which patients are to be randomly assigned to the treatments and the treatments are to be administered double-blinded.

Missing in Exitus

We found that missingness in "Exitus" was likely to have increased when the patients received the combined treatment. If, however, steroidpulse and/or cyclophosphamide were additionally prescribed as the patients' conditions deteriorated, then the observed correlation above might have proven the existence of non-ignorably missing exitus, because the exitus of serious or critically ill patients might have been under-reported precisely because of their conditions. To avert this problem of non-ignorably missing "Exitus", person in charge of collecting and reporting the treatment outcome must be independent of those who administer the treatment and of those who evaluate the prognoses of the patients.

Problems in the Method

There were missing values on both the response and covariates, albeit whose number is small, and they might have been non-ignorably missing, for example, when uncommon treatments are administered. A method for estimating parameters when both the response and the covariates are non-ignorably missing needs to be developed.

References

Agresti, A. (1990) *Categorical data analysis.* New York John Wiley.

Breslow, N.E. and Day, N.E. (1980) *Statistical Methods in Cancer Research.1: The Analysis of Case-Control Studies.* I.A.R.C,Lyon.

Cheigh.J.S, Stenzel.K.H and Rubin.A.L. (1981) *Manual of Clinical Nephrology.* Martinus Nijhoff Publishers.

Freedman, D., Pisani, R., Purres, R. and Adhikri, A. (1991) *Statistics (2nd Edition).* W.W. Norton&Company.

Fujimoto, T. and Dohi, K. (1996) *jin hinyo sikkan.* Nakayama shoten.

Glassock RT. (1996) *Rapidly progressive glomerulonephritis. in The Kidney 5th ed(ed Brenner BM).* Saunders Co. Philadelphia London Toronto Montreal Sydney Tokyo.

Ibrahim, J.G. (1990) Incomplete Data in Generalized Linear Models. *Journal of the American Statistical Association*, 85, 765-769.

Ibrahim, J.G. and Lipsitz, S.R. (1996) Parameter Estimation from Incomplete Data in Binomial Regression When the Missing Data Mechanism is Nonignorable. *Biometrics*, 52, 1071-1078.

Ibrahim, J.G, Lipsitz, S.R. and Chen, M.H. (1999) Missing covariates in generalized linear models when the missing data mechanism is non-

ignorable. *Journal of the Royal Statistical Society B*, 61, 173-190.

Jacob churg. (1982) *Renal Disease.* Igaku-Shoin Tokyo, New York.

Kida, H. and Yoshimura, M. (1996) Shikyutaijinen. *nihon naika gakkai*, 77, 437-440.

Koide, K. and Takahasi, S. (1996) *Manseijinfuzen no yakubutsuryoho.* Tokyo igakusha.

Koyama, A., Igarashi, M., Kobayasi, M., and Members and Coworkers of Research Group on Progressive Renal Diseases. (1997) Natural History and Risk Factors for Immunoglobulin A Nephropathy in Japan. *American Journal of Kidney Diseases* Vol29, No 4, 526-532.

Koyama, T., Ueno, H., Futabin, H., Arimura, Y. and Kida, H. (1997) Rapidly progressive glomerulonephritis. *The research group of progressive renal lesion funded by the Ministry of Health and Welfare.*

Lee, Sang-Gil, and Kanazawa, Y. (1999) Handling "Don't Know" Survey Response: The Case of Japanese Voters on Party Support.

Little, R.J.A and Rubin, D.B. (1987) *Statistical Analysis with Missing Data.* New York John Wiley.

Louis, T.A. (1982) Finding the Observed Information When Using the EM Algorithm. *Journal of the Royal Statistical Society* B, 44, 226-233.

McCullagh, P. (1980) *Generalized Linear Models(2nd Edition).* London: Chapman and Hall.

Michael J D cassidy, Gillian Graskin, John Savill, Charles D Pusey, Andrew J Rees. (1990) Towards a more rapid diagnosis of rapidly progressive glomerulonephritis. *British Medical Journal* 301, August, 329-331.

Miyahara, H., and Tango, T. (1995) *Igakutoukeigaku handbook.* Asakurashoten.

Narukiyo, T. and Asano, Y. (1997) *Kyusokushinkosei-jinen.* Ishiyaku syup-

pan.

Nagase, M. (1995) *Genpatsusei shikyutai shikkan.* Saishin naikagaku taikei, 56, 31-47.

Ota, K. (1997) *jinfuzen chiryogaku.* Nanko do.

Paul R. Rosenbaum. (1995) *Observational Studies.* New York Springer-Verlag.

Robert W.Schrier, M.D. and Carl W. Gottschalk, M.D. (1988) *Diseases of the Kidney.* Little, Brown and Company, Boston/Toronto.

Stewart Cameron, Alex M. Davison, Jean-Pierre Grunfeld, David Kerr, and Eberhard Ritz. (1992) *Oxford Textbook of Clinical Nephrology.* Oxford University Press.

Takeuchi, S. and Kato, E. (1980) *Zusetsu rinsyonaika kouza; 12.* Tokyo, Medical view Co.

Yamaguchi, N. Muro, K. Kikuchi, S. Kobayashi, M. Takahasi, H. and Koyama, T. (1998) Jinfuzen taisaku no genjyo to shorai. *nihon naika gakkai*, 87, 1254-1260.

A Estimation of the Parameters via EM

The ith individual's contribution to the expected log-likelihood is

$$\begin{aligned} &\mathrm{E}_{Y_{mis}}[l(\boldsymbol{\gamma}; \boldsymbol{x}_i, y_i, r_i)|\boldsymbol{x}_i, \boldsymbol{y}_{obs}, r_i] \\ &= \begin{cases} \sum_{y_i=0}^{1} \Pr\{y_i|r_i, \boldsymbol{x}_i, \boldsymbol{\gamma}\} l(\boldsymbol{\gamma}; \boldsymbol{x}_i, y_i, r_i) & \text{if } y_i \text{ is missing} \\ l(\boldsymbol{\gamma}; \boldsymbol{x}_i, y_i, r_i) & \text{if } y_i \text{ is observed,} \end{cases} \end{aligned}$$

where the expectation $\mathrm{E}_{Y_{mis}}$ is taken over the missing data given the observed data. The E-step takes the form of a weighted complete-data log-likelihood

based on $N = \sum_{i=1}^{n} l_i$ observations, where

$$l_i = \begin{cases} 1 & \text{if the } i\text{th individual is observed,} \\ 2 & \text{otherwise.} \end{cases}$$

$\Pr\{y_i|r_i, \boldsymbol{x}_i, \boldsymbol{\gamma}\} = w_{iy_i}^{[t]}$ is the weights corresponding to the missing, and is given by

$$\begin{aligned} w_{iy_i}^{[t]} &= \Pr\{y_i|r_i, \boldsymbol{x}_i, \boldsymbol{\gamma}\} = \frac{\Pr\{y_i, r_i, \boldsymbol{x}_i, \boldsymbol{\gamma}\}}{\Pr\{r_i, \boldsymbol{x}_i, \boldsymbol{\gamma}\}} \\ &= \frac{\Pr\{y_i, r_i, \boldsymbol{x}_i, \boldsymbol{\gamma}\}}{\sum_{y_i=0}^{1} \Pr\{y_i, r_i, \boldsymbol{x}_i, \boldsymbol{\gamma}\}} \\ &= \frac{\Pr\{y_i|\boldsymbol{x}_i, \boldsymbol{\beta}^{(t)}\} \Pr\{r_i|\boldsymbol{z}_i, \boldsymbol{\alpha}^{[t]}\}}{\sum_{y_i=0}^{1} \Pr\{y_i|\boldsymbol{x}_i, \boldsymbol{\beta}^{[t]}\} \Pr\{r_i|\boldsymbol{z}_i, \boldsymbol{\alpha}^{[t]}\}}. \end{aligned} \tag{A.1}$$

If y_i is observed, then $w_{iy_i}^{[t]} = 1$. The E-step of the EM algorithm for all n observations at the $(t+1)$st EM iteration is given by

$$\begin{aligned} Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) &= \sum_{i=1}^{n} \sum_{y_{mis,i}} \Pr\{y_i|r_i, \boldsymbol{x}_i, \boldsymbol{\gamma}\} l(\boldsymbol{\gamma}; \boldsymbol{x}_i, y_i, r_i) \\ &= \sum_{i=1}^{n} \sum_{y_{mis,i}} \{w_{iy_i}^{[t]} l_{y_i}(\boldsymbol{\beta}) + w_{iy_i}^{[t]} l_{r_i|y_i}(\boldsymbol{\alpha})\} \\ &= Q_1(\boldsymbol{\beta}|\boldsymbol{\beta}^{[t]}) + Q_2(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{[t]}), \end{aligned} \tag{A.2}$$

where $\sum_{y_{mis,i}}$ stands for summation taken over when the corresponding response are missing, $\boldsymbol{\alpha}^{[t]}$ and $\boldsymbol{\beta}^{[t]}$ are the parameter estimates at the tth EM iteration, $Q_1(\boldsymbol{\beta}|\boldsymbol{\beta}^{[t]}) = \sum_{i=1}^{n} \sum_{y_{mis,i}} w_{iy_i}^{[t]} l_{y_i}(\boldsymbol{\beta})$ and $Q_2(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{[t]}) = \sum_{i=1}^{n} \sum_{y_{mis,i}} w_{iy_i}^{[t]} l_{r_i|}$ The M-step maximizes the function in (A.2) which is equivalent to doing maximum likelihood estimation. Let $\boldsymbol{\gamma}^{[t](s)}$ denote the estimate of $\boldsymbol{\gamma}$ at the sth iteration of Newton-Raphson within the tth iteration of EM. Then the estimate of $\boldsymbol{\gamma}$ at the $(s+1)$st iteration of the Newton-Raphson procedure within the tth iteration of EM is given by

$$\boldsymbol{\gamma}^{[t](s+1)} = \boldsymbol{\gamma}^{[t](s)} - \left[\ddot{\mathbf{Q}}(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t](s)})\right]^{-1} \dot{\mathbf{Q}}(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t](s)}), \tag{A.3}$$

where $\mathbf{Q}(\gamma|\gamma^{[t](s)}) = \begin{pmatrix} Q_1(\beta|\beta^{[t](s)}) \\ Q_2(\alpha|\alpha^{[t](s)}) \end{pmatrix}$ and $\dot{\mathbf{Q}}(\gamma|\gamma^{[t](s)})$ and $\ddot{\mathbf{Q}}(\gamma|\gamma^{[t](s)})$ are its first and second derivatives with respect to γ respectively. Let $\hat{\gamma}$ denote the estimate of γ at the EM convergence. To obtain the asymptotic covariance matrix of $\hat{\gamma}$, we need the observed information matrix $\mathbf{I}(\gamma)$. We use Louis' formula (1982) to compute the observed information in terms of complete-data quantities. This is given by

$$\begin{aligned}
\mathbf{I}(\gamma) &= E_{Y_{mis}}\left[-\frac{\partial^2 l(\gamma)}{\partial\gamma\partial\gamma^T}\middle|\mathbf{X}, \boldsymbol{y}_{obs}, \boldsymbol{r}\right] \\
&\quad - \left\{E_{Y_{mis}}\left[\mathbf{S}(\gamma;\mathbf{X},\boldsymbol{y},\boldsymbol{r})\mathbf{S}^T(\gamma;\mathbf{X},\boldsymbol{y},\boldsymbol{r})\middle|\mathbf{X},\boldsymbol{y}_{obs},\boldsymbol{r}\right]\right. \\
&\quad - E_{Y_{mis}}\left[\mathbf{S}(\gamma;\mathbf{X},\boldsymbol{y},\boldsymbol{r})|\mathbf{X},\boldsymbol{y}_{obs},\boldsymbol{r}\right] \\
&\quad \left.\times E_{Y_{mis}}\left[\mathbf{S}(\gamma;\mathbf{X},\boldsymbol{y},\boldsymbol{r})|\,\mathbf{X},\boldsymbol{y}_{obs},\boldsymbol{r}\right]^T\right\}
\end{aligned}$$

where $\mathbf{S}(\gamma;\mathbf{X},\boldsymbol{y},\boldsymbol{r}) = \partial l(\gamma)/\partial\gamma = \sum_{i=1}^n \partial l(\gamma;\boldsymbol{x}_i,y_i,r_i)/\partial\gamma$ is the score function. Since $y_1,\ldots,y_n$ are independent observations, the observed information matrix of $\hat{\gamma}$ is

$$\begin{aligned}
&\mathbf{I}(\hat{\gamma}) \\
&= -\sum_{i=1}^n \sum_{y_{mis,i}} \hat{w}_{iy_i} \frac{\partial^2}{\partial\gamma\partial\gamma^T} l(\hat{\gamma};\boldsymbol{x}_i,y_i,r_i) \\
&\quad -\sum_{i=1}^n \sum_{y_{mis,i}} \hat{w}_{iy_i} \mathbf{S}(\hat{\gamma};\boldsymbol{x}_i,y_i,r_i)\cdot\mathbf{S}^T(\hat{\gamma};\boldsymbol{x}_i,y_i,r_i) \\
&\quad +\sum_{i=1}^n \left\{\sum_{y_{mis,i}} \hat{w}_{iy_i}\mathbf{S}(\hat{\gamma};\boldsymbol{x}_i,y_i,r_i)\cdot\sum_{y_{mis,i}} \hat{w}_{iy_i}\mathbf{S}^T(\hat{\gamma};\boldsymbol{x}_i,y_i,r_i)\right\}, \quad \text{(A.4)}
\end{aligned}$$

where $\hat{w}_{iy_i}$ is the estimate of w_{iy_i} at the EM convergence. The inverse of the information matrix in (A.4) is known to be a consistent estimate of the asymptotic variance of $\hat{\gamma}$.

B Missing at random

B.1 Maximum likelihood estimate

The incomplete-data likelihood is for $\gamma = (\beta, \alpha)$ and β and α are distinct,

$$\begin{aligned}
& L_{\boldsymbol{y},\boldsymbol{r}|\mathbf{X}}(\gamma|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,y_i,\ldots,y_m,r_1,\ldots,r_n) \\
&= \prod_{i=1}^{m} L_{y_i,r_i|\boldsymbol{x}_i}(\gamma;\boldsymbol{x}_i,y_i,r_i)\cdot\prod_{i=m+1}^{n}\sum_{y_i=0}^{1} L_{y_i,r_i|\boldsymbol{x}_i}(\gamma;\boldsymbol{x}_i,y_i,r_i) \\
&= \prod_{i=1}^{m} L_{y_i|\boldsymbol{x}_i}(\boldsymbol{\beta};\boldsymbol{x}_i,y_i)\cdot L_{r_i|\boldsymbol{x}_i,y_i}(\boldsymbol{\alpha};\boldsymbol{x}_i,y_i,r_i) \\
&\quad\times\prod_{i=m+1}^{n}\sum_{y_i=0}^{1} L_{y_i|\boldsymbol{x}_i}(\boldsymbol{\beta};\boldsymbol{x}_i,y_i)\cdot L_{r_i|\boldsymbol{x}_i,y_i}(\boldsymbol{\alpha};\boldsymbol{x}_i,y_i,r_i).
\end{aligned}$$

If the data are missing at random (MAR), then the missing data indicator r_i does not depend on y_i and thus $L_{r_i|\boldsymbol{x}_i,y_i}(\boldsymbol{\alpha};\boldsymbol{x}_i,y_i,r_i) = L_{r_i|y_i}(\boldsymbol{\alpha};\boldsymbol{x}_i,r_i)$ and

$$\begin{aligned}
& L_{\boldsymbol{y},\boldsymbol{r}|\mathbf{X}}(\gamma|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,y_i,\ldots,y_m,r_1,\ldots,r_n) \\
&= \prod_{i=1}^{m} L_{y_i|\boldsymbol{x}_i}(\boldsymbol{\beta};\boldsymbol{x}_i,y_i)\cdot L_{r_i|\boldsymbol{x}_i}(\boldsymbol{\alpha};\boldsymbol{x}_i,r_i) \\
&\quad\times\prod_{i=m+1}^{n} L_{r_i|\boldsymbol{x}_i}(\boldsymbol{\alpha};\boldsymbol{x}_i,r_i)\sum_{y_i=0}^{1} L_{y_i|\boldsymbol{x}_i}(\boldsymbol{\beta};\boldsymbol{x}_i,y_i) \\
&= \prod_{i=1}^{m} L_{y_i|\boldsymbol{x}_i}(\boldsymbol{\beta};\boldsymbol{x}_i,y_i)\cdot\prod_{i=1}^{n} L_{r_i|\boldsymbol{x}_i}(\boldsymbol{\alpha};\boldsymbol{x}_i,r_i).
\end{aligned}$$

Therefore, if interest lies in the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$, then the maximum likelihood estimate based on the m completely observed data is equivalent to the maximum likelihood estimate under MAR.

B.2 Observed information matrix

With completely and partially observed parts being written separately in (A.4), then the observed information matrix of $\hat{\gamma}$ under MAR assumption is

$$\mathbf{I}(\hat{\gamma}) = -\sum_{i=1}^{m}\frac{\partial^2}{\partial\gamma\partial\gamma^T}l(\hat{\gamma};\boldsymbol{x}_i,y_i,r_i) - \sum_{i=m+1}^{n}\sum_{y_{mis,i}}\hat{w}_{iy_i}\frac{\partial^2}{\partial\gamma\partial\gamma^T}l(\hat{\gamma};\boldsymbol{x}_i,y_i,r_i)$$

$$
\begin{aligned}
& -\sum_{i=1}^{m} \mathbf{S}(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \cdot \mathbf{S}^T(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \\
& -\sum_{i=m+1}^{n} \sum_{y_{mis,i}} \hat{w}_{iy_i} \mathbf{S}(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \cdot \mathbf{S}^T(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \\
& +\sum_{i=1}^{m} \mathbf{S}(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \cdot \mathbf{S}^T(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \\
& +\sum_{i=m+1}^{n} \left\{ \sum_{y_{mis,i}} \hat{w}_{iy_i} \mathbf{S}(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \cdot \sum_{y_{mis,i}} \hat{w}_{iy_i} \mathbf{S}^T(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \right\} \\
= & -\sum_{i=1}^{m} \frac{\partial^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} l(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \\
& -\sum_{i=m+1}^{n} \left\{ \sum_{y_{mis,i}} \hat{w}_{iy_i} \frac{\partial^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} l(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \right. \\
& \quad + \sum_{y_{mis,i}} \hat{w}_{iy_i} \mathbf{S}(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \cdot \mathbf{S}^T(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \\
& \quad \left. - \sum_{y_{mis,i}} \hat{w}_{iy_i} \mathbf{S}(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \cdot \sum_{y_{mis,i}} \hat{w}_{iy_i} \mathbf{S}^T(\hat{\boldsymbol{\gamma}}; \boldsymbol{x}_i, y_i, r_i) \right\}. \qquad \text{(B.1)}
\end{aligned}
$$

The (j,k)th entry in the second term of (B.1) is

$$
\begin{aligned}
& -\sum_{i=m+1}^{n} \left\{ \sum_{y_{mis,i}} \hat{w}_{iy_i} \hat{\pi}_i (1-\hat{\pi}_i) x_{ij} x_{ik} + \sum_{y_{mis,i}} \hat{w}_{iy_i} (y_i - \hat{\pi}_i) x_{ij} \cdot (y_i - \hat{\pi}_i) x_{i} \right. \\
& \quad \left. - \sum_{y_{mis,i}} \hat{w}_{iy_i} (y_i - \hat{\pi}_i) x_{ij} \cdot \sum_{y_{mis,i}} \hat{w}_{iy_i} (y_i - \hat{\pi}_i) x_{ik} \right\} \\
= & -\sum_{i=m+1}^{n} [-\hat{\pi}_i \hat{\pi}_i (1-\hat{\pi}_i) x_{ij} x_{ik} - (1-\hat{\pi}_i) \hat{\pi}_i (1-\hat{\pi}_i) x_{ij} x_{ik} \\
& \quad + \hat{\pi}_i (1-\hat{\pi}_i)^2 x_{ij} x_{ik} + \hat{\pi}_i^2 (1-\hat{\pi}_i) x_{ij} x_{ik} \\
& \quad - \{\hat{\pi}_i (1-\hat{\pi}_i) x_{ij} - \hat{\pi}_i (1-\hat{\pi}_i) x_{ij}\} \cdot \{\hat{\pi}_i (1-\hat{\pi}_i) x_{ik} - \hat{\pi}_i (1-\hat{\pi}_i) x_{ik} \\
= & \ 0,
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{w}_{iy_i} &= \Pr\{y_i | r_i, \boldsymbol{x}_i, \hat{\boldsymbol{\gamma}}\} = \frac{\Pr\{y_i, r_i, \boldsymbol{x}_i, \hat{\boldsymbol{\gamma}}\}}{\Pr\{r_i, \boldsymbol{x}_i, \hat{\boldsymbol{\gamma}}\}} \\
&= \frac{\Pr\{y_i | \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}\} \Pr\{\boldsymbol{x}_i\} \Pr\{r_i | y_i, \boldsymbol{x}_i, \hat{\boldsymbol{\alpha}}\}}{\sum_{y_i=0}^{1} \Pr\{y_i | \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}\} \Pr\{\boldsymbol{x}_i\} \Pr\{r_i | y_i, \boldsymbol{x}_i, \hat{\boldsymbol{\alpha}}\}}
\end{aligned}
$$

$$= \frac{\Pr\{y_i|\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}\}\Pr\{r_i|\boldsymbol{x}_i,\hat{\boldsymbol{\alpha}}\}}{\sum_{y_i=0}^{1}\Pr\{y_i|\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}\}\Pr\{r_i|\boldsymbol{x}_i,\hat{\boldsymbol{\alpha}}\}}$$
$$= \Pr\{y_i|\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}\},$$

and $\hat{\pi}_i = \Pr\{y_i = 1|\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}\}$.

Therefore the observed information matrix is

$$\mathbf{I}(\hat{\gamma}) = -\sum_{i=1}^{m}\frac{\partial^2}{\partial\hat{\gamma}}l(\hat{\gamma};\boldsymbol{x}_i,y_i,r_i),$$

making the observed information matrix under MAR equivalent to the that with complete case analysis.