LOGISTIC REGRESSION MODELS FOR
MARKERS OF DRUG EFFECTS IN
MEDICATION HISTORIES

BY

HANS PETRI AND T. KAMAKURA

TECHNICAL REPORT NO. 132
AUGUST 1989

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 2 R01 GM21215-15

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
LOGISTIC REGRESSION MODELS FOR MARKERS OF DRUG EFFECTS IN MEDICATION HISTORIES

BY

HANS PETRI AND T. KAMAKURA

TECHNICAL REPORT NO. 132
AUGUST 1989

PREPARED UNDER THE AUSPISES OF
PUBLIC HEALTH SERVICE GRANT 2 R01 GM21215-15

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
LOGISTIC REGRESSION MODELS FOR MARKERS OF DRUG EFFECTS IN MEDICATION HISTORIES

by

Hans Petri
Department of Epidemiology
University of Limburg
6200 MD Maastricht, The Netherlands

and

T. Kamakura
Chuo University
Bunkyo-ku, Tokyo 112, Japan

August 31, 1989
Abstract

Prescription Sequence Analysis (PSA) is based on the observation that a subset of adverse drug reactions are themselves indication for the prescription of another drug. We propose an approach to analyze clustering of prescriptions in medication histories that contain records of a presumably side-effect causing drug $A$ and a side-effect alleviating drug $B$. This set of histories with records of both $A$ and $B$ is analyzed in a logit model that considers the start of $B$ against exposure and non-exposure of $A$. A correction is presented for the periods of $B$ use, in which a new start of $B$ is excluded. A stratification of the history association parameters based on the Akaike Information Criterion is proposed.

Some key words: Adverse drug reaction; Akaike Information Criterion; Logit model; Pharmacy records
1 Introduction

A therapeutic effect of a new drug generally is assessed by means of a clinical trial. Unwanted effects of drugs, however, are much more difficult to detect; only a limited fraction of patients taking a drug experience an adverse reaction, and the serious and fatal reactions generally are rare. However, drugs have been withdrawn for general use after recognition of serious reactions occurring in one in 10,000 to one in 50,000 users. Events with such a low incidence cannot readily be detected in clinical trials, which typically cover 2,000 to 5,000 patients. To exclude with a probability of .95 the occurrence of one or more events in 10,000 people of the general population, the event should not be seen in a sample $3 \times 10,000$ people. This "rule of three" is derived from the Poisson distribution, that yields with these premises a tail probability of $\exp(-3) = .05$ to observe no events while there is at least occurring one in 10,000 in the population.

For the detection of rare adverse drug reactions (ADRs), physicians’ reports on reactions in individual patients have been indispensable. Causality in individual cases, however, is difficult to assess, and adverse reactions that are conditions with a sizeable background incidence in non-users of drugs are generally overlooked. There is a need for methods to detect less obvious relations between drug use and subsequent changes in a patients’ condition. Moreover, the publication of case reports press for further investigation on a population level, and if a perceived problem gets general publicity, society seems to demand an assessment of risk at a short notice.

We will present a method to study a subset of possible adverse drug reactions on a screening basis, within a few weeks. The approach is based on the observation that some unwanted drug effects lead to a condition that is treated with another drug. This concept of Prescription Sequence Analysis (PSA) can be depicted as:

$$drug \ A \rightarrow \ adverse \ reaction \rightarrow \ drug \ B,$$

where $A$ is the drug that causes an adverse drug reaction and $B$ is the drug, or group of drugs, given to treat the condition that followed use of $A$. Patient drug histories should reveal in this situation a relatively high frequency of a dispense of drug $B$ after use of drug $A$ (Petri et al., 1988).

PSA is only applicable where there exists a specific drug therapy to counter an ADR. Therapy should not be to prevent an adverse reaction. Insight into the general use of the drugs should determine if it is suited for PSA: a temporal relation of use of two drugs
may be determined by a third variable, i.e. when the drugs are given for different stages of the same disease. Sometimes additional information, as the indication for use of drugs A and B, will be needed. The analysis should take into account a certain lag-time of the adverse reaction to occur. Primarily, complete and reliable drug histories are needed as basis for the analyses.

2 Medication histories from pharmacy data bases

Drug histories are selected from computerized pharmacy files. About 70% of the Dutch population are insured in the Sickfunds, which requires them to designate one pharmacy to obtain all prescribed drugs. The majority of pharmacies are computerized, and the databases contain virtually complete information on the drugs dispensed on an individual patient level.

A database of all patient drug histories from 5 pharmacies was set up; 65,000 histories spanning a mean period of 31 months (range 10-40 months) are available. The data base begins at the earliest date in each pharmacy at which complete data on dispenses were stored.

3 An application of prescription sequence analysis: use of topical oral antifungal drugs in patients who had inhalational steroids

As an application of PSA we studied one class of drugs with a known side effect, the inhalational steroids for asthma treatment are known to cause oral candidiasis (thrush) in 3-5% of patients (Webb-Johnson, 1977). There is a specific therapy available for oral candidiasis: topical oral forms of drugs that counter the candida infection. The scheme follows:

\[ \text{inhalational steroid} \rightarrow \text{oral candidiasis} \rightarrow \text{antifungal drug} \]

Oral candidiasis is not necessarily followed by antifungal therapy, but the more manifest infections are likely to be. Nystatine and miconazole, the two topical oral antifungal drugs available in the Netherlands are almost exclusively used for candidiasis treatment.

A selection was made of all medication histories that contained records of dispense of inhalational steroids as well as antifungal drugs. The periods of use of inhalational
steroids were estimated from the amount of drug dispensed and the prescribed daily use. A period of 30 days was added to the end of calculated period to adjust for irregular use and pharmacological half-life. For the antifungal drugs, which are typically used for short periods of 1-3 weeks, no period at the end was added, though a repeat prescription was considered to be continuous with the previous one if it was filled within 30 days after the calculated end of use.

4 Data structure and analysis

Patient medication histories were selected from the described data base. The selection criterion was: records of both A) inhalational steroids and B) antifungal medication should be in the history. In these histories the initiation of B medication is to be compared for period before, during, and after use of A, where ”during” is as defined in the previous paragraph. ”Initiation” is the first day of an inferred period of B use. In the comparison the ”before” and ”after” periods will be merged into one category of ”no A use”, as many histories have a period of A use at the beginning or the end of the observation time. Drugs used in the first part of the observation period, but dispensed before the start of the records, will not show up in the patient drug history. Therefore, a window of 90 days was set at the beginning of the observation period, and analysed seperately. As the typical duration of use of one dispense is 90 days, drugs delivered in this period will appear at the redefined start of the medication histories (day 91), whereas before this point no certainty exists if a drug was used. Histories where the single dispense of B was given in the first three months of the observation period will be discarded: in the redefined observation period the condition of use of both A and B is not met.

We want to apply a measure of temporal association of A and B, stratified on the level of the individual histories. A one-level overall measure seems not to be suitable here; The length of use of A and the number of B starts may vary considerabely between the histories. Confounding of the results to a positive association would occur if the number of B starts in the history is correlated to the fraction of time of A use, even if the B starts are distributed uniformly over the whole history, as the example of figure 1 shows: In both patients the incidence of the start of medication with drug B is equal for the periods of A use and non A use, the calculated risk ratio is 1.0 for both patients. A simple addition of the values of the number of B starts and and the number of days of A
use of drug A: xxxx  use of drug B: >>

patient 1

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
>>  >>  >>  >:

patient 2

xxxxxxxxxxxxxxxxxxxxxxxxxxxx
>>  >

Figure 1: Two medication histories with equality of relative risk

use and A non-use of the two patients in a new $2 \times 2$ table would yield a risk ratio that is higher than 1.0. A measure of association of A use and the initiation of B should adjust for the baseline characteristics of the individual histories.

A new period of dispense of a drug can obviously only start after cessation of the previous period of use of the drug. In the analysis of the histories with two selected drugs A and B, this aspect is relevant for the periods of use of B, the marker drug for a possible adverse reaction.

For data sets in which use of B accounts for a substantial fraction of time the periods of B use can be deleted: only the starting day of B is considered. The number of days with no start of B (cells $c_i$ and $d_i$ in Table 2 of second part) is composed from the periods before the start and after the end of use of B: Thus, point '*' is the start of B, and as

Figure 2: Deletion of period of B use in medication history
periods of non-use of \( B \) are considered the intervals 1-2 and 3-4; the interval 2-3 will not be considered in the analysis, if the periods of \( B \) use form a substantial fraction of the total observation period (e.g. > 5%).

5 Results of the data selection

The screening of the data base for evidence of use of drug \( A \) and drug \( B \) was done for the whole reference population (Table 1). Thus, 22 out of 700 users of steroids ever had

Table 1: Selection of medication histories (data source: five pharmacy data bases; duration of records: 21-40 months (mean: 31 months))

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference population</td>
<td>65,000</td>
</tr>
<tr>
<td>users of inhalational steroids ((A))</td>
<td>700</td>
</tr>
<tr>
<td>users of selected antifungal drugs ((B))</td>
<td>815</td>
</tr>
<tr>
<td>users of ( A ) and ( B )</td>
<td>22</td>
</tr>
</tbody>
</table>

a topical antifungal drug. This 3.1% contrasts with the 1.3% (815 out of 65,000) users of antifungal drugs in the reference population. However, as stated earlier, to assess causality we will have to consider the temporal association of use of \( A \) and \( B \). The 3.1% can be seen as a maximum of drug-treated adverse reaction. In the group of 22 users of \( A \) and \( B \) one history was excluded because the only dispense of \( B \) occurred in the first 90 days of the recorded history. A graphical presentatation of the remaining 21 histories is given in figures 6 and 7. For these histories the association of \( A \) use and the initiation of therapy with \( B \) will be analyzed in a logit model. As \( B \) does account only for a small fraction of the observation period, no correction will be made for the periods of \( B \) use (Figure 7 II does not apply for the further text).

6 Logit Model and Common Odds Ratio

In this section we will propose a model to measure the association between use of drug \( A \) and initiation of drug \( B \). A widely used approach is the logit model for analyzing the binary response data set. As our data set is basically binary, we will apply the logit
model. We denote $x_i(t)$ and $y_i(t)$ as two binary variables which represent dispense of drug $A$ and initiation of drug $B$:

\[
\begin{align*}
x_i(t) &= \begin{cases} 
1 & \text{(dispense of drug } A) \\
0 & \text{(otherwise)}, 
\end{cases} \\
y_i(t) &= \begin{cases} 
1 & \text{(the starting day of dispense of drug } B) \\
0 & \text{(otherwise)}. 
\end{cases}
\end{align*}
\]

Here $i$ runs over the set of patient indices: $I$ and $t$ runs over the set of observation days for $i$th patient: $T_i$.

Let $P_i(t)$ be the probability that $y_i(t) = 1$, that is, the probability that $i$-th patient at the time $t$ starts to take drug $B$. The logit model in our data set is expressed by the following probability model:

\[
\log \frac{P_i(t)}{1 - P_i(t)} = \mu_i + \theta x_i(t)
\]  

(1)

Our main concern is to assess the association of the use of drug $A$ and the initiation of drug $B$; the testing problem can be defined as:

\[
H_0 : \theta = 0.
\]

\[
H_1 : \theta > 0.
\]

Table 2: Notation of $2 \times 2$ table( $A$: use of $A$ for the patient $i$, $B$: initiation of $B$ for the patient $i$, $N$: total number of observation days, $a_i$: nr. of days with a start of $B$ during use of $A$, $b_i$: nr. of days with a start of $B$ without use of $A$, $c_i$: nr. of days with no start of $B$ during use of $A$, $d_i$: nr. of days with no start of $B$ without use of $A$ )

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$A$</th>
<th>$B$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_i$</td>
<td>$b_i$</td>
<td>$t_i$</td>
<td>$c_i$</td>
</tr>
<tr>
<td>$B$</td>
<td>$d_i$</td>
<td>$u_i$</td>
<td>$n_i$</td>
<td>$m_i$</td>
</tr>
<tr>
<td>$N_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We can summarize our data in the form of a set of $2 \times 2$ tables for each patient (Table 2 and Table 3). Here the occurrence probabilities of $a_i$ and $b_i$ are $p_{ti}$ and $p_{0i}$ for $i \in I$. 

6
Table 3: The number of days of dispense of drug A and the starting day of dispense of drug B

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>$a_i$</th>
<th>$b_i$</th>
<th>$c_i$</th>
<th>$d_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>0</td>
<td>545</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>24</td>
<td>735</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>130</td>
<td>630</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0</td>
<td>595</td>
<td>163</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>254</td>
<td>506</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>268</td>
<td>866</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>504</td>
<td>630</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>187</td>
<td>947</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>860</td>
<td>274</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>426</td>
<td>708</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>661</td>
<td>473</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>229</td>
<td>905</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>2</td>
<td>45</td>
<td>1088</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>1</td>
<td>22</td>
<td>1112</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>3</td>
<td>364</td>
<td>654</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>181</td>
<td>839</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>1009</td>
<td>11</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>0</td>
<td>728</td>
<td>291</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>0</td>
<td>176</td>
<td>844</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>514</td>
<td>264</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>0</td>
<td>289</td>
<td>489</td>
</tr>
</tbody>
</table>
Assuming a common odds ratio throughout the patient, a parameter of association can be defined:

$$\Psi = \frac{p_{ii}(1 - p_{0i})}{p_{0i}(1 - p_{ii})}.$$  

(2)

To estimate $\Psi$ many authors discuss the problems of unconditional estimation and conditional estimation (Bishop, Fienberg & Holland, 1975; Breslow, 1981). The conditional distribution has good properties (Cox, 1970), but it is difficult to calculate and therefore, the Mantel-Haenszel (1959) estimator is widely used for its simplicity of calculation. This estimator is defined as:

$$\hat{\Psi}_{MH} = \frac{\sum_{i=1}^{K} \frac{a_{ii}}{N_i}}{\sum_{i=1}^{K} \frac{b_{ii}}{N_i}},$$

(3)

where $K$ is $\#\{I\}$. The estimate the of variance of $\hat{\Psi}_{MH}$ was given recently by Robins et al. (1986) and is easy to calculate.

7 Maximum likelihoods estimates of logit model

Assuming the logit model the overall likelihood becomes

$$L = \prod_{i \in I} \prod_{t \in T_i} p_i(t)^{y_i(t)}(1 - p_i(t))^{1 - y_i(t)},$$

(4)

where

$$p_i(t) = \frac{e^{\mu_i + \theta x_i(t)}}{1 + e^{\mu_i + \theta x_i(t)}}.$$  

We will maximize the log-likelihood using the Newton-Raphson method, which needs following derivatives.

$$\frac{\partial \log L}{\partial \theta} = \sum_{i \in I} \sum_{t \in T_i} [y_i(t) - p_i(t)] x_i(t).$$

$$\frac{\partial \log L}{\partial \mu_i} = \sum_{t \in T_i} [y_i(t) - p_i(t)].$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = -\sum_{i \in I} \sum_{t \in T_i} p_i(t) [1 - p_i(t)] x_i(t).$$

$$\frac{\partial^2 \log L}{\partial \mu_i \partial \theta} = -\sum_{t \in T_i} p_i(t) [1 - p_i(t)] x_i(t).$$

$$\frac{\partial^2 \log L}{\partial \mu_i^2} = -\sum_{t \in T_i} p_i(t) [1 - p_i(t)].$$
When we use the Newton-Raphson method we must be careful about selecting the initial values. Expanding the logistic distribution around zero,

\[
\frac{e^x}{1 + e^x} = \frac{1}{2} + \frac{1}{4}x + O(x^3).
\]

Using above formula the following approximates \( p_i(t) \).

\[
p_i(t) = \frac{1}{2} + \frac{1}{4}(\mu_i + \theta x_i(t)).
\]

This approximations give the estimates:

\[
\hat{\theta} = \frac{4 \sum_{i \in I} \left( a_i - \frac{\hat{t}_i}{N_i} \right)}{\sum_{i \in I} \left(1 - \frac{\hat{t}_i}{N_i} \right) n_i},
\]

\[
\hat{\mu}_i = 4 \frac{t_i}{N_i} - \theta \frac{n_i}{N_i} - 2.
\]

Another approximation is to combine multiple strata into one stratum. The likelihood equation is:

\[
a = mp_1.
\]

\[
t = mp_1 + np_0.
\]

Here,

\[
p_0 = \frac{e^\mu}{1 + e^\mu}.
\]

\[
p_1 = \frac{e^{\mu + \theta}}{1 + e^{\mu + \theta}}.
\]

This gives rise to the following estimates:

\[
\tilde{\mu} = \log(b/d). 
\]

\[
\tilde{\theta} = \log(a/c) - \tilde{\mu}.
\]

We note that \( a = \sum_{i=1}^{K} a_i, b = \sum_{i=1}^{K} b_i, c = \sum_{i=1}^{K} c_i \) and \( d = \sum_{i=1}^{K} d_i \). The initial value of \((\tilde{\theta}, \tilde{\mu}, \ldots, \tilde{\mu})\) can be one of the initial value sets. Table 4 shows the combined one of 21 patients.

Here we obtain the parameter estimates which are shown in Table 5. We note that in this logit model the stratified ML estimate of the odds ratio is given by

\[
\hat{\Psi} = e^{\hat{\theta}}
\]

\[
= e^{0.3388} = 1.3991.
\]
Table 4: The combined $2 \times 2$ table of 21 patients

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>A</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>16</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>B</td>
<td>8011</td>
<td>12451</td>
<td>20462</td>
</tr>
<tr>
<td></td>
<td>8027</td>
<td>12466</td>
<td>20493</td>
</tr>
</tbody>
</table>

Table 5: The estimates of parameters of the logit models for the 21 patients

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Normal Deviate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>0.3358</td>
<td>0.7103</td>
<td>0.2388</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
</tr>
<tr>
<td>$\mu_2$</td>
</tr>
<tr>
<td>$\mu_3$</td>
</tr>
<tr>
<td>$\mu_4$</td>
</tr>
<tr>
<td>$\mu_5$</td>
</tr>
<tr>
<td>$\mu_6$</td>
</tr>
<tr>
<td>$\mu_7$</td>
</tr>
<tr>
<td>$\mu_8$</td>
</tr>
<tr>
<td>$\mu_9$</td>
</tr>
<tr>
<td>$\mu_{10}$</td>
</tr>
<tr>
<td>$\mu_{11}$</td>
</tr>
<tr>
<td>$\mu_{12}$</td>
</tr>
<tr>
<td>$\mu_{13}$</td>
</tr>
<tr>
<td>$\mu_{14}$</td>
</tr>
<tr>
<td>$\mu_{15}$</td>
</tr>
<tr>
<td>$\mu_{16}$</td>
</tr>
<tr>
<td>$\mu_{17}$</td>
</tr>
<tr>
<td>$\mu_{18}$</td>
</tr>
<tr>
<td>$\mu_{19}$</td>
</tr>
<tr>
<td>$\mu_{20}$</td>
</tr>
<tr>
<td>$\mu_{21}$</td>
</tr>
</tbody>
</table>
The Mantel-Haenszel estimates of $\Psi$ is

$$\hat{\Psi}_{MH} = 1.4817$$

and the its variance can be estimated based on the paper Robins et al. (1986).

$$Var(\hat{\Psi}) = 0.6809.$$ 

Thus testing of the association results in the following calculation:

$$\frac{1.4817 - 1}{\sqrt{0.6809}} = 0.5838.$$ 

The P-value is here 0.2797. The Mantel-Haenszel estimate is somewhat larger and gives a larger P-value in comparison with the logit model.

Secondly we will calculate the likelihood ratio test. The log-likelihood ratio of the $\theta$ value becomes:

$$-2[\log L(0, \hat{\mu}_1, \ldots, \hat{\mu}_{21}) - \log L(\hat{\theta}, \hat{\mu}_1, \ldots, \hat{\mu}_{21})] = -2[(-226.142) - (-225.335)]$$

$$= 1.614.$$ 

This likelihood ratio test denotes a rather weak positive association for the case of 21 strata.

When we consider the one-stratum approach, the following likelihood ratio result comes out:

$$-2[\log L(0, \hat{\mu}) - \log L(\hat{\theta}, \hat{\mu})] = -2[(-233.040) - (-231.303)]$$

$$= 3.474 < \chi^2_{0.95}(1) = 3.842.$$ 

In this case the P-value of the likelihood ratio statistic is 0.0623. The score statistic and the associated P-value are 0.5055 and 0.0799 respectively. Both score and likelihood ratio tests indicate a moderate association in the one-stratum model.

We will now consider the problem of selection of models with different number of strata. From the foregoing results we can anticipate that stratification may greatly change the degree of association. A one-stratum model results in a moderately positive association, but this model is not realistic, as we pointed out at the end of the chapter on data structure and analysis. Thus we need to investigate the best stratification from
the point of view of model selection. Here we use the \( AIC \) – concept developed by Akaike(1974) which is defined by,

\[
AIC = -2(\text{Maximum Log} - \text{likelihood}) + 2(\text{the number of parameters included in the model}).
\]

We consider firstly the next two models;

(I) \text{One stratum model}

(II) \text{Full(21) strata model}

The two \( AIC \) values are calculated as follows:

\[
AIC(I) = -2(-231.303) + 2 \times 2 = 466.607.
\]

\[
AIC(II) = -2(-225.335) + 2 \times 22 = 494.670.
\]

Based on the \( AIC \) values Model I would be selected.

We can also use the likelihood ratio test (\( H_0: \mu_1 = \ldots, = \mu_{21} (= \mu_0) \)).

\[
-2[\log L(\tilde{\theta}, \tilde{\mu}_0) - \log L(\hat{\theta}, \hat{\mu}_1, \ldots, \hat{\mu}_{21})] = -2[(-231.303) - (-225.335)]
\]

\[
= 11.936 < \chi^2_{0.95}(20) = 31.41.
\]

Here we cannot reject the null hypothesis.

8 Discussion

In this paper we considered the measure of association of use of drug \( A \) and initiation of drug \( B \) in the logistic model. The logistic approach makes sense in that multiple maxima of the likelihood function are unlikely unless there are either very limited data or gross discrepancies with model exists (Cox, 1970; Everitt, 1987). This is confirmed by calculating the second derivatives of the log-likelihood, a Hessian matrix, which is proved to be a nonpositive definite matrix and yields a stronger result, with at most one maximum. Furthermore we can prove that our model which treats the independent strata does not have multiple maxima even in the case the strata contain no occurrence of \( a_i \) or \( b_i \). The odds ratio estimate from the logistic model is somewhat smaller than the Mantel-Haenszel estimate. The Mantel-Haenszel estimate is less appropriate for our data set because of the occurrence of a zero value in one cell of each of the 21 \( 2 \times 2 \) tables.
(Rothman, 1986, pp. 195–196). Furthermore, we can treat the problem of stratification by using the likelihood approach. It is important to find the best stratification because stratification may have a great influence on the behavior of the estimates of the common parameter $\theta$.

![Graph showing two-sided error bands of estimates of $\mu_i$'s of 21 patients with one standard error.](image)

**Figure 3:** Two-sided error bands of the estimates of $\mu_i$'s of 21 patients with one standard error.

In our analysis we treated only two models, but there may be more possibilities. If we calculate all possibilities, the number of situations becomes the number of all possible partitions of set $I$, and the fast algorithm which finds effectively the best stratification based on some criteria, like AIC is required and left to more investigation. Here we conclude this section just presenting a model which consists of the smaller number of strata and gives the smaller AIC value from the graphical technique. The absolute value of 21 estimates of $\mu_i$ with one standard error and a half standard error are shown in Fig. 3 and Fig. 4, where the $\hat{\mu}_i$'s are estimated from the full 21-strata model. These figures suggest us the stratification, $\{1\}, \{2,4,15\}, \{3,5,13,18\}, \{6,7,8,9,10,11,12,14,16,17,19,20,21\}$. The AIC value, the estimate of $\theta$ and its P-value is 462.63, 0.3208 and 0.1790, respectively. We note here that the AIC-stratified model does not raise the value of the estimate of the association parameter $\theta$, but that its advantage lies in the smaller P-value.
Figure 4: Two-sided error bands of the estimates of $\mu_i$’s of 21 patients with a half standard error

9 Acknowledgement

We are grateful to Professor Bradley Efron at Stanford University for helpful discussions.

References


A Appendix 1

We will consider here for one person the initiation of drug therapy in a time period $T_i$ as a point process, with an exclusion for renewal during the $k$ periods of use of the drug. $T_{nu} = T_1 + T_2 + ... + T_m$ is the total of periods of non-use of the drug in $T_i$. If for a person the probability to start medication is constant for each day, the number of starts can be considered to be the outcome of a Poisson process; conditional on the number of starts $k$, the occurrence of the $k$ starts can be considered to have a uniform distribution over the period $T_{nu}$, $T_{nu}$ being composed from the successive periods of non-use, separated by the starting points: The actual data should show existing deviations from a uniform distribution. In the logit model of prescription sequence analysis a deviation from uniformity of the distribution the starts of a side-effect causing drug would have no consequences for the analysis: an association between the use of drug $A$ and the start of the side-effect alleviating drug $B$ will be detected regardless of the distribution of drug $A$ use over time. Figure 5

B Appendix 2

**Theorem 1** The usual logit model in which the probability of occurrence of a event is defined by

$$p_i = \frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)}$$
LEGEND:  
Ti: total observation period for one person  
*: start of period of drug use  
————: period of drug use  
T1-T4: periods of non-use of drug  
Tnu: total period of non-use of drug  

Figure 5: Distribution of the initiation of drug therapy

has the log-likelihood whose Hessian matrix is nonpositive definite.

<Proof> The rs-th element of the Hessian matrix is

\[ H_{rs} = \frac{\partial^2 \log L}{\partial \beta_r \beta_s} = -\sum_{i=1}^{n} x_{ir} x_{is} p_i (1 - p_i). \]

Here we set \( y_{ir} = x_{ir} \sqrt{p_i (1 - p_i)} \), \( y'_r = (y_{1r}, ..., y_{nr}) \) and \( Y = (y_1, ..., y_p) \). Then \(-H = Y'Y\) and it is clear H is nonpositive definite.

C Appendix 3: proof of the convexity of the stratified logit likelihood

Lemma 1 A symmetric \( n \times n \) matrix \( A \) is positive definite iff all principal minors \( |A_k| (k = 1, ..., n) \) are positive.

Lemma 2 Let $A$ be a positive definite $n \times n$ matrix, and $B$ be the $(n + 1) \times (n + 1)$ matrix,

$$B = \begin{pmatrix} A & b \\ b' & \alpha \end{pmatrix}.$$

$B$ is positive definite iff $|B| > 0$.


Theorem 2 If $n_1, n_2, \ldots, n_s$ and $m_1, m_2, \ldots, m_s$ in the $2 \times 2$ tables are not zeros, then, the Hessian matrix of the log-likelihood of the stratified logit model is negative definite.

<Proof> We define $u_i$ and $v_i$ as follows for simplicity:

$$u_i = \left\{ \frac{e^{u_i + \theta}}{1 + e^{u_i + \theta}} \right\} n_i,$$

$$v_i = \left\{ \frac{e^{u_i}}{1 + e^{u_i}} \right\} m_i.$$

We express the Hessian matrix by $H$ and $-H$ by $V$. It is clear that $|V_1|$ (the upper-left minor) is positive. Suppose $|V_{s+1}| > 0$ for all $\ell \leq s$, then we can show that $|V_{s+1}| > 0$ using Lemma 2. By induction $|V_{s+1}| > 0$ for all $\ell \leq K$ and $V$ is positive definite from Lemma 1.

$$|V_{s+1}| = \begin{bmatrix} (u_s + v_s) - (u_s, 0, \ldots, 0)V_s^{-1} & 0 \\ 0 & \vdots \\ 0 & 0 \end{bmatrix} |V_s|.$$

Here $V_s^{11}$ is calculated as follows:

$$V_s^{11} = \left( \sum_{i=1}^{K} u_i - \sum_{j=1}^{s-1} \frac{u^2_j}{u_j + v_j} \right)^{-1}.$$

Therefore,

$$|V_{s+1}| = \begin{bmatrix} (u_s + v_s) - \frac{u^2_s}{\sum_{i=1}^{K} u_i - \sum_{j=1}^{s-1} \frac{u^2_j}{u_j + v_j}} \end{bmatrix} |V_s|$$

$$\geq \frac{(u_s + v_s) \sum_{i=1}^{K} u_i - u^2_s}{\sum_{i=1}^{K} u_i - \sum_{j=1}^{s-1} \frac{u^2_j}{u_j + v_j}} |V_s|$$

$$= \frac{v_s \sum_{i=1}^{K} u_i + u_s \sum_{i=s+1}^{K} u_i}{\sum_{i=1}^{K} u_i - \sum_{j=1}^{s-1} \frac{u^2_j}{u_j + v_j}} |V_s| > 0$$

17
Here we note that the denominator is

$$\sum_{i=1}^{K} u_i - \sum_{j=1}^{s-1} \frac{u_j^2}{u_j + v_j} > \sum_{i=s}^{K} u_i > 0.$$
Figure 6. Periods of Steroid and Anti-fungal Use in 21 Patients

1 Year

Steroid
Anti-fungal
Figure 7. I. Structure of the Medication Histories
   II. Deletion of Periods of Drug B Use

I

Record Begins

PSA Begins

Record Ends

II

Period Deleted in the Analysis

Steriod

Anti-fungal