Research Report
Department of Statistics
Göteborg University
Sweden

Prediction of work resumption among men and women with lower back- and neck pain in a Swedish population

Anders Persson

Research Report 2002:4
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Dept of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se/stat
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

PREDICTION OF WORK RESUMPTION AMONG MEN AND WOMEN WITH LOWER BACK- AND NECK PAIN IN A SWEDISH POPULATION

by Anders Persson

Department of Statistics, Göteborg University, Sweden

ABSTRACT

An approach based on Bayes theorem is used to predict the binary outcome of work resumption *X*, where $X=1$ if no work resumption and $X=0$ otherwise, given a vector of discrete predictors $\mathbf{Z}$ for men and women with lower back- and neck pain in a Swedish population. In this application the predictors have a complex dependency structure. Hierarchical cluster analysis is used to create independent groups of dependent predictors such that predictors within groups are dependent while predictors in different groups are independent. The main purpose is to estimate the probability $P(X=1|\mathbf{z})$ and to calculate confidence intervals for this probability. Based on these estimates one may decide whether a given person should be predicted as healthy or as non-healthy, and predictive values are calculated in order to evaluate of the performance of the prediction analysis. The results are compared with the frequently used ordinary logistic regression method without interactions. It is found that ignoring the correlations between the predictors may give seriously misleading results. Also, the problem with missing values is discussed.

Key words: Confidence intervals; Hierarchical cluster analysis; Logistic regression; Prediction; Predictive value; Work resumption.

1 INTRODUCTION

In many applications the aim is to predict a binary outcome given the value of a set of predictor variables. A commonly used method for this situation is ordinary logistic regression (Cox (1970); Hosmer and Lemeshow (1989); McCullagh and Nelder (1989) and Neter et al. (1996)). In many applications, the predictors have a complex dependency structure, which might be difficult to capture with the logistic model. Although the use of interaction terms works well in a logistic model with few predictors, problems may arise when there are many predictors. The reason for this is that there is a total of $2^k - 1$ β-parameters to estimate if all interactions are included. For obvious reasons it is almost impossible to include all interactions if there are many predictors.

In this paper we apply a method suggested by Jonsson and Persson (2002) which is based on Bayes theorem to predict the outcome variable 'work resumption' $(X = 0)$ and 'no work resumption' $(X = 1)$ among men and women with back- and neck pain diagnosis, conditional on the values of a discrete vector of predictors **Z**.

This paper is motivated by the fact that the number of long-term sick-listed individuals has been increasing persistently in Sweden and in many other countries. Back- and neck pain is one of the most frequently cases behind long-term sick-listing (Bergendorff et al. (1997) and Hansson and Hansson (1999)). Since the middle of the 80s the National Social Insurance Board (RFV) has conducted studies to identify important factors affecting health state improvement and work resumption. Due to increased efforts on economic and personal resources, including interventions to improve the propensity of work resumptions, it has resulted in amount of positive changes since the beginning of the 90s (Riksförsäkringsverket (1995) and Persson and Tasiran (2001)). But, during the period 1997-2001 the numbers of individuals who have been sick-listed longer than 365 days have increased from 75,000 to 120,000. The relative increase during the 4-year period have been about 30% per year and the number of earlier retirements/temporary disability pensions have increased from 423,000 to 450,000. Including waiting period, sick pay, sickness allowance and earlier retirements/temporary disability pensions it corresponds to 800,000 full-time annual jobs or 14 percent of the population at the ages

18-65. The associated costs for health insurance are 108 billion SEK according to the state budget 2002 (SOU (2002)).

A sample consisting of 1575 full-time working employed was available for the analyses. Four sub-groups were of special interest: men and women with back- or neck pain diagnosis, and were treated separately. The individuals in the sample were followed-up during a 2-year period and predictions were possible at 90 days, 1 year and 2 years after sick-listing, respectively.

The process of prediction proceeded in the following two stages. In a first step, the probabilities $\pi = P(X=1|\mathbf{z})$ were estimated and confidence intervals were calculated for each probability. In a second step, new subjects were sampled sequentially from the same population by simulations to make predictions of 'no work resumption' given the values of a set of predictors based on the estimates in the first step.

The plan of the paper is as follows. Section 2 starts with a brief description of the material. In Section 3 the statistical methods are described. Section 4 deals with estimation of $P(X=1|\mathbf{z})$ and calculation of confidence intervals for these probabilities. This section ends with prediction of 'no work resumption', and presents measures for prediction ability such as predictive values. Finally, in Section 5, some concluding remarks are given.

2 MATERIAL

A sample of 1575 full-time working employed sick-listed for at least 28 days because of lower back- or neck pain diagnosis followed-up during a two-year period was available for the analyses. Data were collected by the National Social Insurance Board (RFV) sequentially during the period November 1994 until October 1995 represented by 5 different counties of Sweden; Stockholm, Kristianstad, Västmanland, Västernorrland and Göteborg. Three time points were of special interest: 90 days, 1 year and 2 years after sick-listing. Individuals with both lower back- and neck pain diagnosis (240) were excluded from the analyses due to difficulties with confounding effects. Each of these 4 sub-groups (by sex and diagnosis) was treated separately due to large differences between their patterns of work resumption. For a detailed description of the material see Bergendorff et al. (1997); Bergendorff et al. (2001); Riksförsäkringsverket och Sahlgrenska universitetssjukhuset (1997) and Hansson and Hansson (1999).

Unfortunately, the data quality was rather low since there were considerable amounts of missing values on some predictor variables (see discussion in Section 4.1). Furthermore, only 5 counties participated in the study. Hence, the results were not representative for the whole population of Sweden.

Sometimes the term 'healthy' and 'non-healthy' will be used for simplicity rather than 'work resumption' and 'no work resumption', respectively. The state 'healthy' was defined as a sick-listed person who has become able to work. A person, who was fully or partially sick-listed, early retirement or entitled to temporary disability pension, was defined as a 'non-healthy' person (Bergendorff et al. (2001)). Occasionally, we use the abbreviation MB90, MB1Y, MN90, MN1Y, WB90, WB1Y, WN90 and WN1Y, where M=men, W=women, B=back pain, N=neck pain. 90=90 days and 1Y=1 year.

Baseline characteristics. *Sex*, *Age* (Z_1), *Diagnosis* and *County*. There were a total of 883 females (56%) and 692 males at the ages 18-59. The mean(SD) age was 42(10) years for all groups. In the analyses, Age was dichotomized where Age=1 if a person was older than 31 years and 0 if a person was younger than 31 years. High age was a positive factor for 'no work resumption' in all groups except for women with lower back pain diagnosis.

Table 1 below shows the prevalence in the sub-groups at 90 days, 1 year and 2 year after sick-listing. People with lower back pain recovered faster than those with neck pain. Among persons with lower back pain there were 42 percent healthy within 90 days, 79 percent within 1 year and 87 percent within 2 years. The corresponding figures for people with neck pain were 39, 73 and 81 percent, respectively. Men with lower back pain recovered faster than women with the same diagnosis, while there was no significant difference between men and women with neck pain (Bergendorff et al. (2001)).

	90 days	1 year	2 years
Men/Back	0.54	0.17	0.11
Men/Neck	0.60	0.30	0.21
Women/Back	0.63	0.24	0.15
Women/Neck	0.63	0.25	0.18

Table 1: Prevalence's at 90 days, 1 year and 2 years after sick-listing.

There was a strong connection between sex and diagnosis. Table 2 shows that men suffered more frequently from back problems (79%) as compared to women (63%), while women suffered more frequently from neck problems (37%) as compared to the men (21%). The diagnoses varied between the counties in the material. Table 3 below shows the distribution of lower back- and neck diagnosis in the 5 counties. Lower back pain was the most frequent cause of sick-listing in Stockholm (73%) while neck pain was most frequent in Västmanland (36%).

County	**Men/Back**	**Men/Neck**	**Women/Back**	**Women/Neck**
Stockholm	169	41	154	76
Kristianstad	87	30	88	56
Västmanland	64	23	74	56
Västernorrland	76	23	104	52
Göteborg	149	30	132	91
Total	**545**	**147**	**552**	**331**

Table 2: Number of cases of sick-listing by county, sex and diagnosis.

County	**Back (%)**	**Neck (%)**
Stockholm	**73**	27
Kristianstad	67	33
Västmanland	64	**36**
Västernorrland	71	29
Göteborg	70	30

Table 3: Distribution of lower back- and neck pain diagnosis in the 5 counties.

Socioeconomic factors. *Education* (Z_2), *Ethnicity* (Z_3) and *Household income* (Z_4). Education was defined on a 3 level ordinal scale with 1 as lowest and 3 as highest degree of education. Level 1 and 2 representing low education (=1) and level 3=high education (=0). Ethnicity is a 20 level nominal variable where 1=Swedish and 2-20 representing non-Swedish (=0). Finally, Household income was a continuous variable ranging from 900 to 175,000 SEK, dichotomized as 1 if > 7000 SEK and 0 otherwise.

Psychical working environment. *Demand* (Z_5), *Control* (Z_6), *Strain* (Z_7) and *Attitude* (Z_8). Demand was expressed as self experienced demands on their place of work, scaled 25 (low)-100 (high), where 25-70 was defined as low (=0) and 70-100 as high (=1). Control is the possibility of affecting their own working environment scaled 25 (low)-100 (high), where 25-70 was defined as low (=1) and 70-100 as high (=0). Strain is simply the ratio between Demand and Control, where 0.25-0.84 was defined as low (=0) and ≥ 0.84 as high (=1). Attitude was measured on a scale 3 (low)-9 (high) where 0-4 was defined as low (=0) and ≥ 5 as high (=1).

Physical working environment. *Inconvenient working environment* (Z_9), *Heavy lifts* (Z_{10}) and *Suitable working tasks* (Z_{11}). By 'Inconvenient working environment' and 'Heavy lifts' we mean 4 level variable ranging from 1 (yes, often) to 4 (no, never), where 1-2 was defined as yes (=1) and 3-4 as no (=0). Finally, by 'Suitable working tasks' is meant that the employer was willing to adjust the working tasks in agreement with the individual's state of health, where 1=no and 0=yes.

Family and social networks. *Sick-listing in the family* (Z_{12}), *Temporary disability pension/early retirement in the family* (Z_{13}) and *Offered temporary disability pension/early retirement* (Z_{14}). All variables were dichotomous where 1=yes and 0=no.

Health state. *Work ability* (Z_{15}), *Comorbidity* (Z_{16}) and *Smoking* (Z_{17}). Working ability was subjectively assessed on a scale ranking from 1 (low) to 10 (high), where 1-4 was defined as bad working ability and 5-10 as good working ability. By Comorbidity we mean that the individual has other diseases than lower back- or neck pain, where 1-2 was defined as no (=0) and 3 as yes (=1). Smoking was a 3 level variable defined as yes or never smoked (=1) and quit smoking (=0).

Administrative interventions. The presence of *Complete rehabilitation plan* (Z_{18}) was a dichotomous variable defined as 1=yes and 0=no.

Predictor	**MB (%)** $n = 545$	**MN (%)** $n = 147$	**WB (%)** $n = 552$	**WN (%)** $n = 331$	**p-value**
Age	84	82	86	83	.35
Education	90	94	93	95	.11
Ethnicity	19	21	14	26	<.01
Household income	96	99	91	91	<.01
Demand	48	63	56	66	<.01
Control	34	32	48	56	<.01
Strain	62	74	73	83	<.01
Attitude	84	83	90	87	.07
Inconvenient working environment	85	93	85	89	.17
Heavy lifts	82	82	80	76	.48
Suitable working tasks	50	64	62	63	.02
Sick-listing in the family	13	13	9	8	.15
TDP/ER in the family	17	15	14	12	.31
Offered TDP/ER	11	24	12	17	.01
Work ability	51	50	53	45	.26
Comorbidity	8	15	8	8	.13
Smoking	67	64	74	75	.05
Rehabilitation plan	18	21	26	24	.02

Table 4: Descriptive statistics and χ^2-test of equal proportions between the 4 sub-groups. The proportions in the table are given that all predictors equals to 1 (see definitions in Table A1 in Appendix). The abbreviation TDP/ER denotes Temporary Disability Pension/Early Retirement.

3 STATISTICAL METHODS

A method based on Bayes theorem for predicting a binary outcome $X = 0,1$ given the values of a vector of discrete predictors **Z**, suggested by Jonsson and Persson (2002) is used for the analyses. The probability π was estimated according to (3) and 95% confidence limits according to (12:(*i*)) in the latter work. At baseline i.e. after 28 days of sick-listing a large set of predictors was available from the material. In a previous study (Bergendorff et al. (2001)) a list of potential predictors has been proposed for prediction of work resumption among men and women with lower back- and neck pain (see Table

5). The predictors were chosen on basis of probability plots. In a second step, a hierarchical clustering method (Anderberg (1973) and Jobson (1992)) have been used to create independent groups of dependent predictors both given $X = 1$ and $X = 0$. That is, for a given value of X the purpose is to identify groups of predictors such that predictors within groups are dependent but at the same time are independent of predictors in other groups. Consequently, it is not necessarily the same predictors in the groups given $X = 1$ and $X = 0$, respectively. In addition to the cluster analysis Pearson's correlation coefficient have been calculated between the predictors both given $X = 1$ and $X = 0$ to examine the dependency structure in detail. Although a χ^2-test of independence in a 2×2 contingency table may be sufficient, the correlation coefficient is perhaps a better descriptive measure of association between the predictors. In fact, the χ^2-test and Pearson's correlation coefficient are related by $r = \{n^{-1}X^2\}^{1/2}$, where r is the correlation coefficient, X^2 is the value of the chi-square statistic and n is the number of observations.

	Men				**Women**			
	Back		**Neck**		**Back**		**Neck**	
Predictor	**90d**	**1y**	**90d**	**1y**	**90d**	**1y**	**90d**	**1y**
Z_1	X	X	X	X				
Z_2		X	(X)	X				X
Z_3	X	(X)						
Z_4							X	X
Z_5	X	(X)	(X)	(X)	X	X	X	(X)
Z_6			X					
Z_7			(X)				X	
Z_8			(X)				(X)	
Z_9	(X)	(X)					(X)	(X)
Z_{10}	(X)					(X)		(X)
Z_{11}	X	X		(X)			(X)	(X)
Z_{12}	X							
Z_{13}			(X)	X				
Z_{14}			X	(X)		(X)	(X)	(X)
Z_{15}	X	X	X	X	X	X	X	X
Z_{16}	X	X	X	X	X	X	X	X
Z_{17}			(X)	X				
Z_{18}	X	(X)	(X)	(X)	X	X	X	X

Table 5: Potential predictors for prediction analysis at 90 days and 1 year after sick-listing (Bergendorff et al. (2001)). Predictors marked with (X) were not included in the models. See Table A1 in Appendix for labels to the predictors.

One possibility to test whether a predictor has a significant effect on the outcome is a stepwise logistic regression. However, this method cannot be used for testing whether the predictors are dependent or not conditionally on $X = 1$ and $X = 0$. This follows easily from the illustrations in (Jonsson and Persson (2002), p. 7). Furthermore, with 8 predictors there are up to 255 β–coefficients to be tested in a pre-test, and this give rise to inferential problems. But, there is another possibility that we might consider. Let $\mathbf{z}_1 = (z_i = 1, \mathbf{z}_r)$ be the vector of all predictors with the constraint that the *i*th predictor takes on the value 1 and $\mathbf{z}_0 = (z_i = 0, \mathbf{z}_r)$ that the *i*th predictor takes on the value 0, where $\mathbf{z}_r$ is a subset of $\mathbf{z}$ when the *i*th predictor is excluded. The effect of the predictor Z_i given $\mathbf{z}_r$ can be expressed as the estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$, where $\hat{\pi}_1 = \hat{P}(X = 1 | \mathbf{z}_1)$ and $\hat{\pi}_0 = \hat{P}(X = 1 | \mathbf{z}_0)$. For example, if $\mathbf{z} = (z_1, z_2)$ and $\mathbf{z}_r = (z_2)$ then $\mathbf{z}_1 = (z_1 = 1, z_2)$ and $\mathbf{z}_0 = (z_1 = 0, z_2)$. Since z_2 can take on the values 0 or 1 there are 2 possible outcomes for $\hat{\delta}$, $\hat{\pi}_1$ and $\hat{\pi}_0$, respectively. The difference δ is estimable if and only if there are observations on both z_1 and z_2. Let n' be the number of estimable δ's. Then, $\max\{n'\} = 2^{k-1}$, where *k* is the number of predictors. We want to test the hypothesis $H_0 : \delta = 0$ given that the predictors $\mathbf{z}_r$ are in the model, against the alternative $H_A : \delta \neq 0$. It can be performed in many ways. With few estimable δ's a Sign test may be appropriate. If the number of observations is sufficiently large and normal distribution of the $\hat{\delta}$'s's can be assumed, a test based on normality may be better, or if the distribution is at least symmetric a Wilcoxon Signed Rank test may be appropriate (Altman (1991)). These tests require a large set of predictors and very few missing values. For example, if there are two predictors in the model, there are only two differences to calculate. This will be further explained in Section 4.1, and examples will be given in Table 9.

4 PREDICTION

This section is devoted to prediction of the binary outcome X = 'work resumption' conditional on the values of a vector of discrete predictors **Z**. We are primarily interested in predicting 'no work resumption' $(X = 1)$. The reason for this is that among non-healthy persons it was desirable to find characteristics such that appropriate interventions e.g.

rehabilitation actions that gain work resumption can be taken as soon as possible after sick-listing. Predictions were made 90 days and 1 year after sick-listing, respectively. A detailed discussion is given in Section 4.1 for men with back pain (90 days) only. But, in Section 4.2 we summarize and compare the prediction results from the remaining sub-groups as well.

4.1 Men with Lower Back Pain (90 days)

There were 545 men with lower back pain diagnosis available for the analysis. Initially, there were 10 potential predictors of interest (see Table 5), but these have been reduced to 8 predictors. There were considerable amounts of missing values for most of the predictors (see Table A1 in Appendix). For example, the predictor 'Suitable working tasks' had 333 (61%) missing values. With k binary predictors there are 2^k possible outcomes for π. For example, with 8 predictors there are 256 various outcomes that require a rather large sample size and few missing values. The sample size needed for estimation of the π's depend on the distribution of the cell frequencies (Jonsson and Persson (2002)).

Table 6 shows the dependency structure among the 8 chosen predictors given 'no work resumption' $(X=1)$ and 'work resumption' $(X=0)$, respectively. Note that there were not the same predictors in the groups given $X=1$ and given $X=0$. That is, the composition of predictors across groups affecting the probability of 'no work resumption' is different from the probability of 'work resumption'. The following dependency structure was obtained from the hierarchical cluster analysis. Note that it is not the same predictors in Table 6 as in the simulation example in Section 5.1 in Jonsson and Persson (2002).

Group	**Predictors associated with 'no work resumption'** $(X=1)$
1	Rehabilitation plan (Z_{18}), Demand (Z_5), Suitable working tasks (Z_{11})
2	Sick-listing in the family (Z_{12}), Ethnicity (Z_3)
3	Comorbidity (Z_{16}), Work ability (Z_{15}) Age (Z_1)

Group	**Predictors associated with 'work resumption'** $(X=0)$
1	Comorbidity (Z_{16}), Demand (Z_5), Suitable working tasks (Z_{11}), Ethnicity (Z_3)
2	Rehabilitation plan (Z_{18}), Work ability (Z_{15}) Age (Z_1), Sick-listing in the family (Z_{12})

Table 6: Result of hierarchical cluster analysis for men with lower back pain (90 days).

Table 7 and 8 shows the correlations between pairs of predictors given $X=1$ and $X=0$. It is seen that the hierarchical clustering method to some extent agrees with the correlation coefficients between pairs. But, from Table 7 it is seen that Age (Z_1) in group 3 given $X=1$ is pairwise independent of Work ability (Z_{15}) and Comorbidity (Z_{16}) with correlations .00 and -.05, respectively. However, Age (Z_1) is at the same time independent of every predictor in the group 1 and 2. Furthermore, in group 2 given $X=0$, Table 8 shows that Rehabilitation plan (Z_{18}) is pairwise independent of Age (Z_1), Sick-listing in the family (Z_{12}) and Work ability (Z_{15}) with correlations .01, .08 and .06, respectively. But, Rehabilitation plan (Z_{18}) is at the same time independent of every predictor in group 1. It should be noticed that pairwise independency is not the same as simultaneously independency.

	Z_1	Z_3	Z_5	Z_9	Z_{10}	Z_{11}	Z_{12}	Z_{15}	Z_{16}	Z_{18}
Z_1	1	-	-	-	-	-	-	-	-	-
Z_3	.07	1	-	-	-	-	-	-	-	-
Z_5	-.04	.11	1	-	-	-	-	-	-	-
Z_9	-.05	.13	.09	1	-	-	-	-	-	-
Z_{10}	.06	.05	.10	**.48**	1	-	-	-	-	-
Z_{11}	.07	-.01	**.36**	**.27**	**.33**	1	-	-	-	-
Z_{12}	.03	**-.15**	-.04	**-.16**	.01	-.12	1	-	-	-
Z_{15}	-.05	.10	.14	.08	-.05	**.25**	.01	1	-	-
Z_{16}	.00	**.23**	.05	.02	-.05	.09	-.03	**.15**	1	-
Z_{18}	.01	-.03	**.19**	**.16**	.14	**.25**	.01	**.14**	.12	1

Table 7: Correlation matrix for predictors among men with lower back pain diagnosis (90 days) associated with 'no work resumption' $(X=1)$. Significant correlations (5%) are marked with bold type $\left(n^{(1)}=295\right)$.

	Z_1	Z_3	Z_5	Z_9	Z_{10}	Z_{11}	Z_{12}	Z_{15}	Z_{16}	Z_{18}
Z_1	1	-	-	-	-	-	-	-	-	-
Z_3	.15	1	-	-	-	-	-	-	-	-
Z_5	.06	.09	1	-	-	-	-	-	-	-
Z_9	-.04	-.06	.07	1	-	-	-	-	-	-
Z_{10}	-.15	-.17	.05	**.44**	1	-	-	-	-	-
Z_{11}	-.02	.12	**.23**	**.26**	.17	1	-	-	-	-
Z_{12}	**-.20**	.05	.07	.11	.04	**.23**	1	-	-	-
Z_{15}	-.09	.10	.01	.06	-.02	.07	**.17**	1	-	-
Z_{16}	.08	**.29**	.14	-.01	.02	**.19**	.07	**.20**	1	-
Z_{18}	.01	-.08	.06	.10	.12	.01	.08	**.23**	-.04	1

Table 8: Correlation matrix for predictors among men with lower back pain diagnosis (90 days) associated with 'work resumption' $(X=0)$. Significant correlations (5%) are marked with bold type $(n^{(0)}=250)$.

Sparse contingency tables often contain cells having zero frequency counts or missing values. Cells for which a nonzero count is impossible because of the design of the study are sometimes referred to as structural zeros. In this application, however, we are only concerned with missing values and sampling zeros i.e. nonzero counts are possible, but a zero occurs because of random variation. Sampling zeros are especially likely to arise when the sample is small and the contingency table has many cells (Agresti (1991)). Out of the 545 observations there were 186 observations available for prediction and only 50 (20%) of the 256 probabilities were estimable due to missing values and sampling zeros. It means that if new individuals are sampled from the same population in the same way as in the original survey, it is likely that some individuals have values of the predictors such that predictions for those subjects are not possible. The numbers of missing values for each predictor are presented in Table A1 in Appendix.

The separate effect for each predictor in the model is illustrated in Table 9 with the δ-test (see also Figures 9-16 for plots of the $\hat{\delta}$'s for each predictor). The results in Table 9 show that individuals with complete rehabilitation plan, bad work ability, sick-listing in the family and people who did not have suitable working tasks had higher probability of 'no work resumption'. But, Age, Comorbidity, Demand and Ethnicity did not show any significant differences indicating that these should be excluded from the model. However, due to the fact that there are very few estimable δ's (n') the reliability of the test result may be questionable. The reason for that n' is relatively small compared to the 50 estimable probabilities is that the test require values of every predictor in the vector $\mathbf{z}_r$ both given $Z_i = 1$ and $Z_i = 0$. Otherwise, δ is not estimable for that combination of **Z**. With 8 predictors the maximal value of n' is 128. None of the predictors in Table 9 has a value of n' greater than 20 and the value for Comorbidity is as low as 7.

Predictor	n'	**Mean($\hat{\delta}$)**	**Median($\hat{\delta}$)**	**Std. dev($\hat{\delta}$)**	**p-value**
Age (Z_1)	9	.037	.008	.102	.16
Rehab. plan (Z_{18})	11	.384	.291	.234	<.01
Comorbidity (Z_{16})	7	.018	.060	.202	.94
Work ability (Z_{15})	17	.175	.223	.199	<.01
Demand (Z_5)	17	-.041	-.064	.096	.09
Sick-listing in the family (Z_{12})	10	.243	.241	.200	<.01
Suitable working tasks (Z_{11})	18	.194	.178	.117	<.01
Ethnicity (Z_3)	10	.093	.114	.221	.19

Table 9: Descriptive statistics and a Sign test of the differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for testing if the predictors have an effect on the outcome variable.

Table 10 shows the frequencies of predicted work resumption versus the true state.

		True state		
		Healthy	Non-healthy	
Predicted State	Healthy	81	33	114
	Non-healthy	15	57	72
		96	90	186

Table 10: Predicted and true state of work resumption for men with lower back pain 90 days. Out of the 545 individuals only 186 observations were available for prediction due to missing values on the predictor variables.

In order to evaluate the prediction ability, simulations have been used to sample new individuals (100,000) from the same population. The prediction ability was evaluated by predictive values, relative predictive values and proportion of correct classifications. But the predictive value could in some cases be misleading without reference to the prevalence. For example, a predictive value of .92 and prevalence .90 is obviously not as good as if the prevalence was .20, say. Therefore, it seems more reasonable to use relative predictive values. The latter show the relative gain in predicting the outcome rather than simply guessing the outcome in accordance with the prevalence. From Table

1 it can be seen that the prevalence at 90 days was 0.54. The predictive value for 'no work resumption' and 'work resumption' was 0.76 and 0.74, respectively, with the corresponding relative predictive values 42% and 37%. The proportion of correct classifications was 0.75 (see also Table 12 for comparisons with the remaining sub-groups).

We recall that various values of $\mathbf{Z}$ give different values of π. Figure 1 shows the ordered values of $\hat{\pi}$ and associated confidence limits. Since, it is not possible from the figure to identify the values of the predictors represented by the index variable on the x-axis, Table A2 in Appendix presents all estimable π's, confidence limits and $\hat{V}[\hat{\pi}]$. The following examples illustrate how to interpret the results.

Example 1. Mr. A is a Swedish man older than 31 years entitled to rehabilitation plan. He has other diseases than lower back pain, bad working ability and he experience high demand at his place of work, where his working tasks are not appropriate for him. Furthermore, he has no cases of sick-listing in his family. Mr. A has a probability of 'no work resumption' equal to 0.97 with confidence limits (0.38; 0.99).

Example 2. Mr. B is a Swedish man older than 31 years with no rehabilitation plan. He has no other diseases than lower back pain, good working ability and he does not experience high demand at his place of work, where his working tasks are appropriate for him. Furthermore, he has no cases of sick-listing in his family. Mr. B has a probability of 'no work resumption' equal to 0.14 with confidence limits (0.09; 0.26).

4.2 Comparison of the Prediction Results for All Sub-groups

Table 11 shows the proportion of $\hat{\pi} \leq \frac{1}{2}$ and $\hat{\pi} > \frac{1}{2}$ in all the 4 sub-groups. It is seen that the proportion of $\hat{\pi} > \frac{1}{2}$ is rather high at 90 days for all groups and low after 1 year. For a detailed examination of the estimated probabilities, see Figures 1-8 and Table A2, which show the ordered values of $\hat{\pi}$ and associated confidence limits.

The results obtained from the prediction analysis for all sub-groups are showed in Table 11 and 12. It is seen from Table 11 that the proportion of estimable probabilities is rather low fore some groups and high for others. But, there is different number of predictors in the sub-groups. Of course, it easier to obtain a higher proportion of estimable probabilities with fewer predictors.

Group	n	No. of obs. for prediction	No. of predictors	No. of probabilities to estimate	No. of estimable probabilities	Proportion of $\hat{\pi} > \frac{1}{2}$
MB90	545	186	8	256	50	0.58
MB1Y	545	161	6	64	23	0.17
MN90	147	47	6	64	9	0.56
MN1Y	147	80	6	64	21	0.24
WB90	552	309	4	16	16	0.87
WB1Y	552	303	5	32	26	0.42
WN90	331	111	6	64	16	0.56
WN1Y	331	138	5	32	14	0.36

Table 11: Basic statistics for all sub-groups, separately.

In Table 12 below it can be seen that the prediction ability after 1 year is better performed as compared to 90 days. But, there are no differences in prediction ability between men and women and between lower back and neck pain diagnosis.

Group	Prevalence	Predictive value non-healthy	Relative predictive value non-healthy	Predictive value healthy	Relative predictive value healthy	Proportion of correct classified
MB90	0.54	0.76	42%	0.74	37%	0.75
MB1Y	0.17	0.59	244%	0.91	435%	0.86
MN90	0.60	0.68	14%	0.82	37%	0.70
MN1Y	0.30	0.68	127%	0.81	170%	0.78
WB90	0.63	0.81	28%	0.63	0%	0.73
WB1Y	0.24	0.65	169%	0.86	258%	0.82
WN90	0.63	0.81	29%	0.62	-2%	0.73
WN1Y	0.25	0.70	199%	0.85	240%	0.82

Table 12: Prediction ability for all sub-groups, separately.

5 DISCUSSION

This paper is an application of an approach suggested by Jonsson and Persson (2002) based on Bayes theorem. The aim is to predict the outcome 'no work resumption' conditionally on the values of a set of discrete predictors, and also to make CI statements. It is emphasized that problems may arise when some of the predictors have a considerable amount of missing values. The consequences of getting missing values may be serious. First, the number of observations available for prediction and the number of estimable probabilities decreases. Secondly, if new subjects are sampled sequentially from the same population, it is likely that we obtain individuals with values on the vector of predictors such that π is not estimable. The material in this application contains many missing values, which would have justified the use of fewer predictors in the model. In the latter case we would have obtained relatively more observations for prediction. Fewer predictors do not necessarily alter the prediction ability. Also, the number of parameters to estimate increases dramatically as the number of predictors in the model increases.

The proposed method works very well for most cases. Correct specification of the dependency structure is a matter of crucial importance. The assumption of independent predictors when they in fact are correlated may lead to seriously misleading results concerning bias and variance (Jonsson and Persson (2002)). For example, for men with lower back pain (90 days) where $Z_1 = Z_3 = Z_5 = Z_{12} = Z_{18} = 0$ and $Z_{11} = Z_{15} = Z_{16} = 1$ we obtain $\hat{\pi} = 0.76$ for the Bayes approach with a corresponding estimate of 0.29 for an ordinary logistic regression model. In this paper we have used the decision rule; if $\pi > \frac{1}{2}$ then a given subject is predicted 'no work resumption' and if $\pi \leq \frac{1}{2}$ then the subject is predicted 'work resumption'. The choice of the limit is somewhat arbitrary, but in a real life situation the estimated probability will be used in conjunction with other sources of information about the sick-listed person to reach a decision whether e.g. interventions should be taken.

A test for detecting separate variable effects in the model was suggested. Since the test depends on the number of predictors in the model it seems inappropriate to use such a test for materials with many missing value and few potential predictor variables.

The results of the predictions showed that the prediction ability after 1 year was better performed as compared to 90 days, as measured by relative predictive values. But, there were no differences in prediction ability between men and women and between lower back- and neck pain diagnosis.

ACKNOWLEDGMENTS

I would like to acknowledge a special appreciation to my supervisor Ph. D. Robert Jonsson for advice and encouraging support during the work of this paper. Thanks to Christian Sonesson for constructive comments and suggestions on earlier versions of the manuscript.

REFERENCES

[1] Agresti, A. (1990) *Categorical Data Analysis*. New York: John Wiley & Sons.

[2] Altman, D.G. (1991) *Practical Statistics for Medical Research*. London: Chapman and Hall.

[3] Anderberg, M.R. (1973) *Cluster Analysis for Applications*. New York: Academic Press.

[4] Bergendorff, S. Hansson, E., Hansson, T. and Jonsson, R. (2001) (In Swedish) *Vad kan förutsäga utfallet av en sjukskrivning?* Rygg och Nacke 8. Stockholm: Riksförsäkringsverket och Sahlgrenska universitetssjukhuset.

[5] Cox, D.R. (1970) *Analysis of Binary Data*. London: Chapman and Hall.

[6] Hosmer, D.W. and Lemeshow, S. (1989) *Applied Logistic Regression*. New York: Wiley.

[7] Jobson, J.D. (1992) *Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods*. New York: Springer-Verlag.

[8] Jonsson, R. and Persson, A. (2002) *Bayes Prediction of Binary Outcomes Based on Correlated Discrete Predictors*. Research report 2002:3, Department of Statistics, Göteborg University.

[9] McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models* (2nd ed.). London: Chapman and Hall.

[10] Neter et al. (1996) *Applied Linear Regression Models* (3rd ed.). London: Irwin.

[11] Persson, A. and Tasiran, A.C. (2001) *Analysis of Spatial Effects on Long-term Sicknesses*. Seminar paper 2001:5, Department of Statistics, Göteborg University.

[12] Riksförsäkringsverket (1995) (In Swedish) *RIKS-LS - en undersökning om långvarig sjukskrivning och rehabilitering*. RFV REDOVISAR 1995:20. Stockholm: Riksförsäkringsverket.

[13] Bergendorff, S. Hansson, E., Hansson, T., Palmer, E., Westin, M. and Zetterberg, C. (1997) (In Swedish) *Projektbeskrivning och undersökningsgrupp*. Rygg och Nacke 1. Stockholm: Riksförsäkringsverket och Sahlgrenska universitetssjukhuset.

[14] Riksförsäkringsverket och Sahlgrenska universitetssjukhuset (1997) (In Swedish) *Enkäter till undersökningsgruppen och försäkringskassan*. Rygg och Nacke 2. Stockholm: Riksförsäkringsverket och Sahlgrenska universitetssjukhuset.

[15] Hansson, E. and Hansson, T. (1999) (In Swedish) *Medicinsk åtgärder för sjukskrivna för rygg- och nackbesvär*. Rygg och Nacke 3. Stockholm: Riksförsäkringsverket och Sahlgrenska Universitetssjukhuset.

[16] Riksförsäkringsverket (2000) *Sjukfrånvaro och förtidspension*, RFV ANALYSERAR 2000:2. Stockholm: Riksförsäkringsverket (in Swedish).

[17] SOU (2002) (In Swedish) *Handlingsplan för ökad hälsa i arbetslivet*. Statens Offentliga Utredningar, 2002:2, Stockholm: Fritzes.

APPENDIX

Predictor	Label	No. of missing values (%)
Z_1	Age $\begin{cases} 1 \text{ if } \geq 31 \text{ years} \\ 0 \text{ if } < 31 \text{ years} \end{cases}$	0(0)
Z_2	Education $\begin{cases} 1 \text{ if low} \\ 0 \text{ if high} \end{cases}$	208(38)
Z_3	Ethnicity $\begin{cases} 1 \text{ if Non-swedish} \\ 0 \text{ if Swedish} \end{cases}$	170(31)
Z_4	Household income $\begin{cases} 1 \text{ if } \geq 7000 \text{ SEK} \\ 0 \text{ if } < 7000 \text{ SEK} \end{cases}$	239(44)
Z_5	Demand $\begin{cases} 1 \text{ if high} \\ 0 \text{ if low} \end{cases}$	267(49)
Z_6	Control $\begin{cases} 1 \text{ if low} \\ 0 \text{ if high} \end{cases}$	267(49)
Z_7	Strain $\begin{cases} 1 \text{ if high} \\ 0 \text{ if low} \end{cases}$	267(49)
Z_8	Attitude $\begin{cases} 1 \text{ if high moral} \\ 0 \text{ if low moral} \end{cases}$	180(33)
Z_9	Inconvenient working environment $\begin{cases} 1 \text{ if yes} \\ 0 \text{ if no} \end{cases}$	263(48)
Z_{10}	Heavy lifts $\begin{cases} 1 \text{ if yes} \\ 0 \text{ if no} \end{cases}$	265(49)
Z_{11}	Suitable working tasks $\begin{cases} 1 \text{ if no} \\ 0 \text{ if yes} \end{cases}$	333(61)
Z_{12}	Sick-listing in the family $\begin{cases} 1 \text{ if yes} \\ 0 \text{ if no} \end{cases}$	187(34)
Z_{13}	TDP/ER in the family $\begin{cases} 1 \text{ if yes} \\ 0 \text{ if no} \end{cases}$	181(33)
Z_{14}	Offered TDP/ER $\begin{cases} 1 \text{ if yes} \\ 0 \text{ if no} \end{cases}$	260(48)
Z_{15}	Work ability $\begin{cases} 1 \text{ if bad } (\leq 4) \\ 0 \text{ if good } (> 4) \end{cases}$	178(33)
Z_{16}	Comorbidity $\begin{cases} 1 \text{ if yes} \\ 0 \text{ if no} \end{cases}$	173(32)
Z_{17}	Smoking $\begin{cases} 1 \text{ if yes or never} \\ 0 \text{ if quited} \end{cases}$	184(34)
Z_{18}	Rehabilitation plan $\begin{cases} 1 \text{ if yes} \\ 0 \text{ if no} \end{cases}$	7(1)

Table A1: Labels to predictors and the number of missing values (%) for MB$(n = 545)$. The abbreviation TDP/ER denotes Temporary Disability Pension/Early Retirement.

Z_1	Z_3	Z_5	Z_{11}	Z_{12}	Z_{15}	Z_{16}	Z_{18}	$\hat{\pi}$	$\hat{\pi}_{Lower}$	$\hat{\pi}_{Upper}$	$\hat{V}[\hat{\pi}]$
1	1	1	0	0	0	1	0	0.03	0.01	0.61	0.001
1	0	1	0	0	0	0	0	0.08	0.04	0.24	0.001
0	0	1	0	0	0	0	0	0.08	0.04	0.38	0.002
1	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.14**	**0.09**	**0.26**	**0.001**
0	0	0	0	0	0	0	0	0.14	0.07	0.41	0.005
0	0	0	0	1	1	0	1	0.18	0.06	0.68	0.029
1	0	0	1	0	0	0	0	0.19	0.11	0.38	0.004
0	0	0	1	0	0	0	0	0.20	0.09	0.50	0.010
1	1	1	0	0	0	0	0	0.23	0.09	0.61	0.021
1	0	1	1	0	0	0	0	0.23	0.15	0.40	0.004
1	1	1	1	0	0	0	0	0.30	0.15	0.58	0.015
0	0	1	0	0	1	0	0	0.30	0.14	0.61	0.020
1	0	1	0	0	1	0	0	0.31	0.18	0.54	0.010
1	1	1	1	1	1	1	0	0.37	0.12	0.78	0.080
1	1	0	0	0	0	0	0	0.39	0.19	0.67	0.024
1	0	1	0	1	0	0	0	0.41	0.19	0.71	0.033
1	1	1	1	0	1	1	0	0.44	0.19	0.74	0.040
0	0	0	0	0	1	0	0	0.46	0.25	0.70	0.021
0	0	0	1	1	1	0	1	0.46	0.16	0.81	0.084
1	0	1	0	0	1	0	1	0.47	0.19	0.78	0.051
1	0	0	0	0	1	0	0	0.47	0.33	0.62	0.007
0	0	1	1	1	1	0	0	0.51	0.22	0.79	0.045
0	0	0	1	0	1	0	0	0.55	0.29	0.77	0.025
1	0	1	0	1	1	0	0	0.55	0.27	0.79	0.034
1	0	0	1	0	1	0	0	0.56	0.37	0.72	0.011
1	0	0	0	0	1	0	1	0.57	0.29	0.79	0.030
1	0	0	0	1	0	0	0	0.57	0.30	0.79	0.027
0	1	1	0	0	1	0	0	0.60	0.24	0.85	0.058
0	0	1	1	0	1	0	0	0.61	0.34	0.80	0.021
1	0	1	1	0	1	0	0	0.62	0.43	0.75	0.008
1	0	0	1	1	0	0	0	0.66	0.34	0.85	0.027
1	0	1	0	0	0	0	1	0.66	0.22	0.89	0.082
1	1	1	1	0	1	0	0	0.69	0.41	0.85	0.016
1	0	0	0	1	1	0	0	0.71	0.40	0.86	0.019
1	0	1	1	1	0	0	0	0.71	0.38	0.87	0.021
1	0	0	0	0	0	0	1	0.74	0.29	0.92	0.046
1	1	1	0	0	1	0	1	0.75	0.28	0.92	0.049
1	0	0	1	0	1	1	0	0.76	0.30	0.92	0.040
0	0	0	1	0	1	1	0	0.76	0.26	0.93	0.060
1	1	0	0	0	1	0	0	0.77	0.45	0.89	0.014
1	0	0	1	1	1	0	0	0.78	0.43	0.90	0.016
1	0	1	1	1	1	0	0	0.81	0.47	0.92	0.011
1	0	1	1	0	1	1	0	0.83	0.35	0.94	0.024
1	1	1	1	0	1	1	1	0.83	0.37	0.94	0.021
1	0	0	1	0	1	0	1	0.84	0.50	0.93	0.009
1	0	1	1	0	1	0	1	0.91	0.62	0.96	0.003
1	0	0	1	0	0	0	1	0.92	0.39	0.97	0.007
1	1	1	1	0	1	0	1	0.93	0.54	0.97	0.003
1	0	1	1	0	0	0	1	0.96	0.43	0.99	0.002
1	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**0.97**	**0.38**	**0.99**	**0.002**

Table A2: Ordered predicted values and associated CI's for various combinations of **Z** in accordance with Figure 1 (MB90). The variance corresponds to formula (8) in Jonsson and Persson (2002). See also Table A1 for labels to the predictors.

LEGENDS TO FIGURES

Figure 1: The probability of 'no work resumption' and associated CI's (MB90).

Figure 2: The probability of 'no work resumption' and associated CI's (MB1Y).

Figure 3: The probability of 'no work resumption' and associated CI's (MN90).

Figure 4: The probability of 'no work resumption' and associated CI's (MN1Y).

Figure 5: The probability of 'no work resumption' and associated CI's (WB90).

Figure 6: The probability of 'no work resumption' and associated CI's (WB1Y).

Figure 7: The probability of 'no work resumption' and associated CI's (WN90).

Figure 8: The probability of 'no work resumption' and associated CI's (WN1Y).

Figure 9: Estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for Age (MB90).

Figure 10: Estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for Rehabilitation plan (MB90).

Figure 11: Estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for Comorbidity (MB90).

Figure 12: Estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for Work ability (MB90).

Figure 13: Estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for Demand (MB90).

Figure 14: Estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for Sick-listing in the family (MB90).

Figure 15: Estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for Suitable working tasks (MB90).

Figure 16: Estimated differences $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ for Ethnicity (MB90).

Figure 1

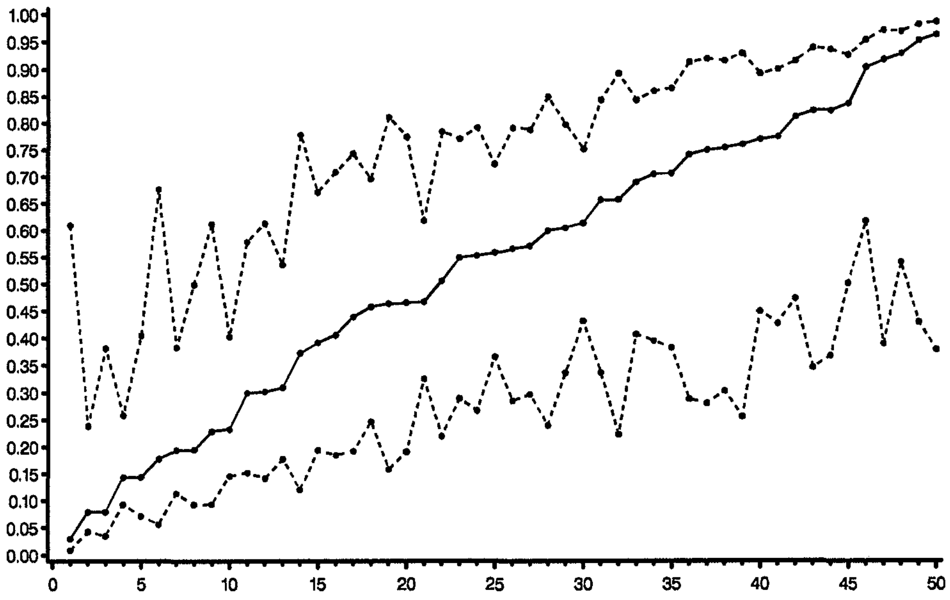


Figure 2

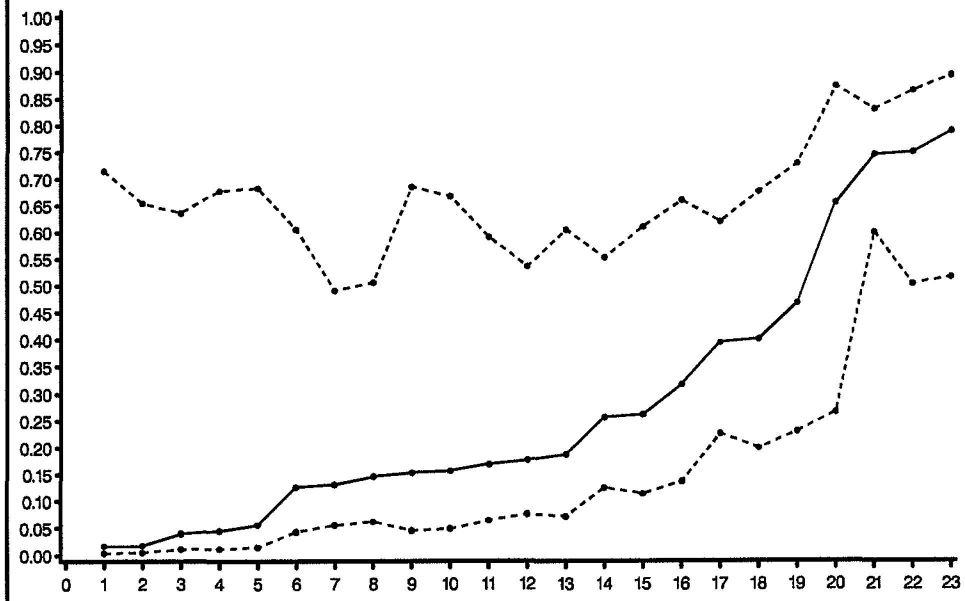


Figure 3

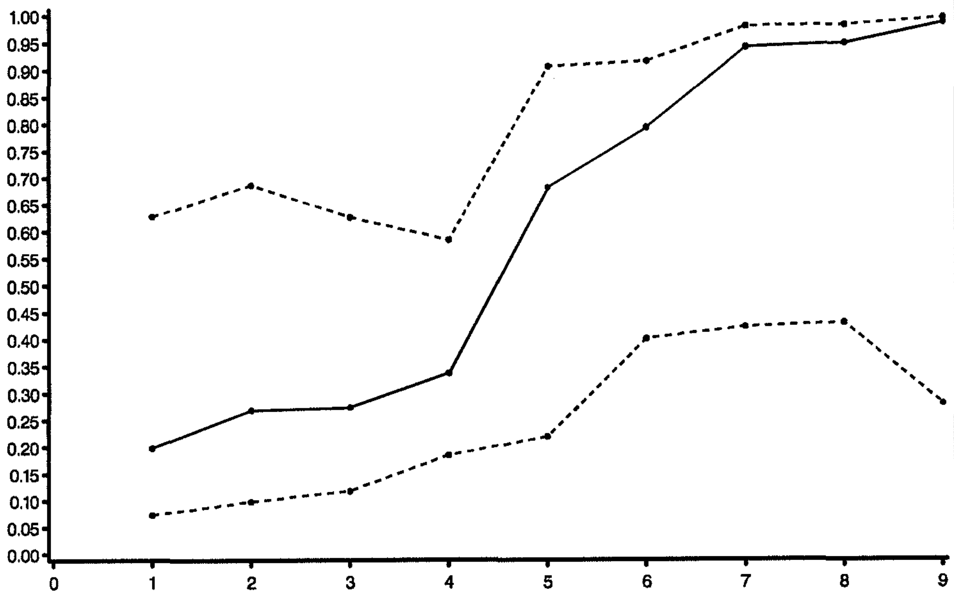


Figure 4

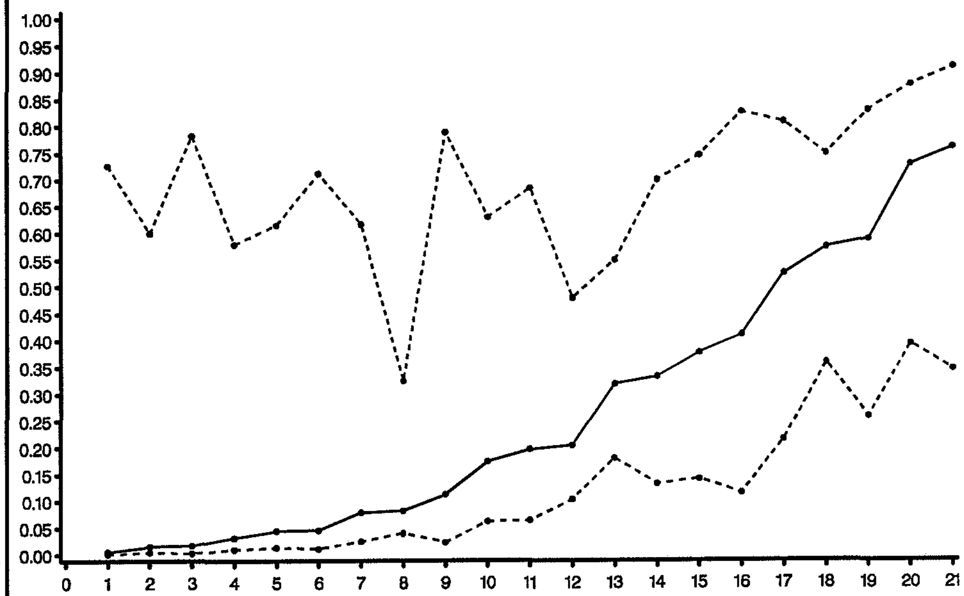


Figure 5

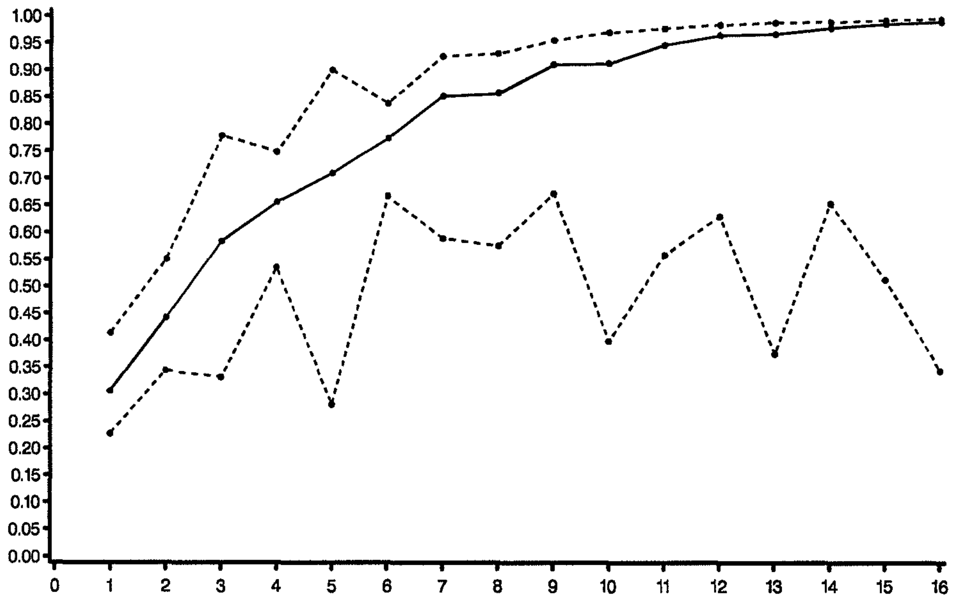


Figure 6

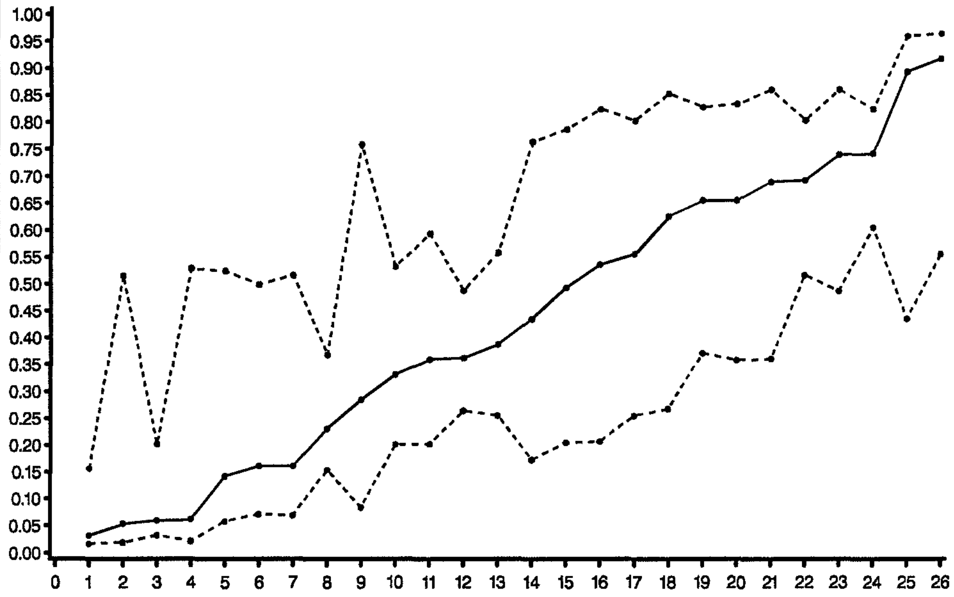


Figure 7

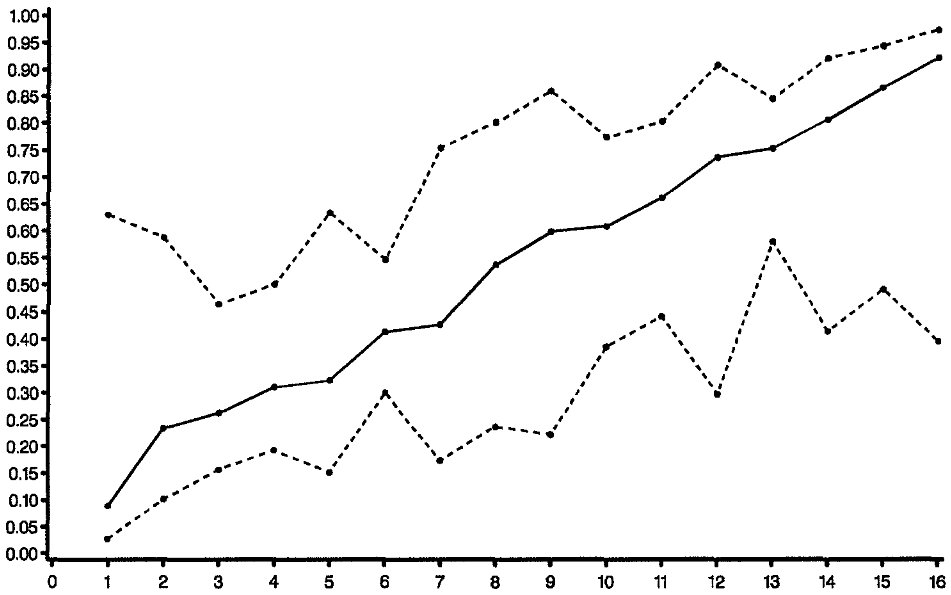


Figure 8

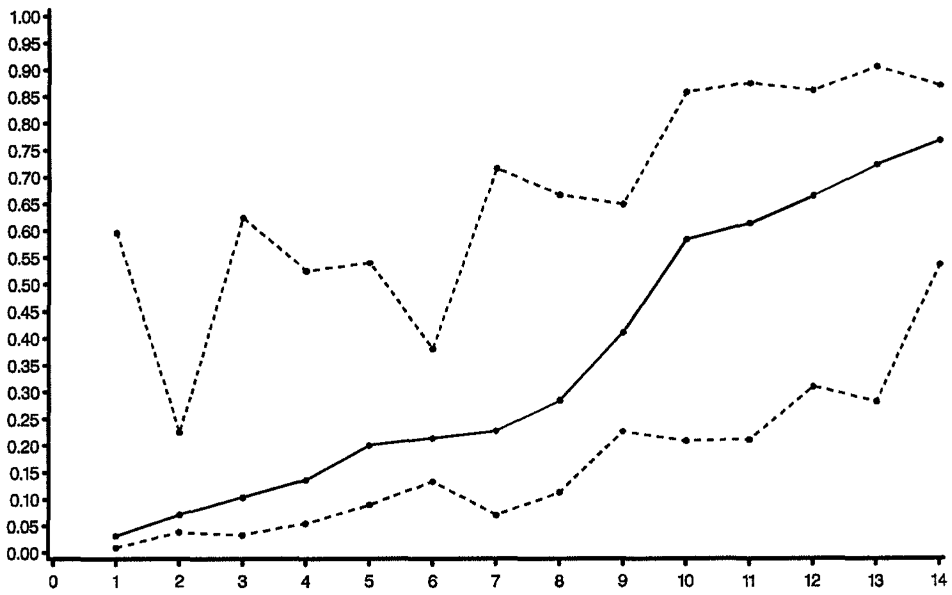


Figure 9

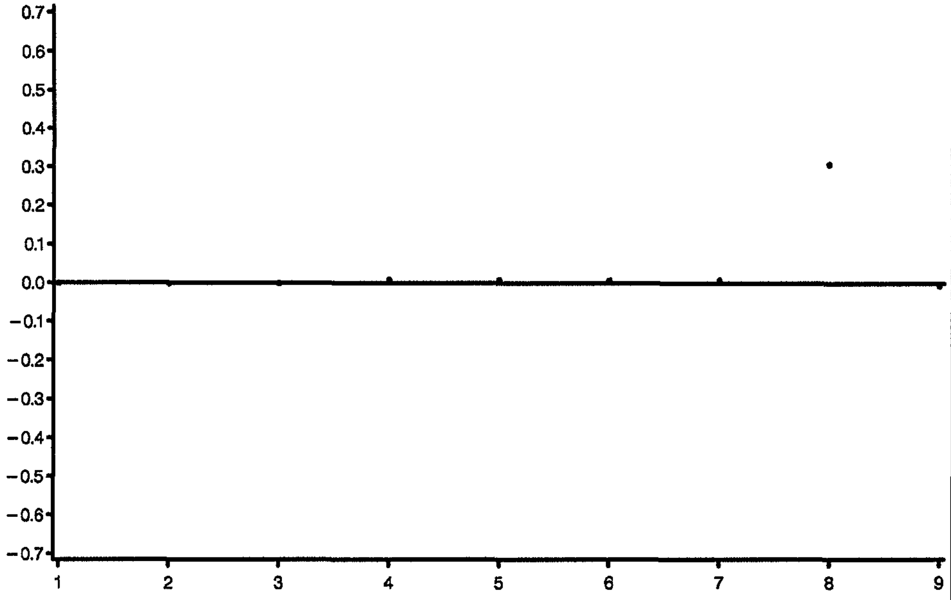


Figure 10

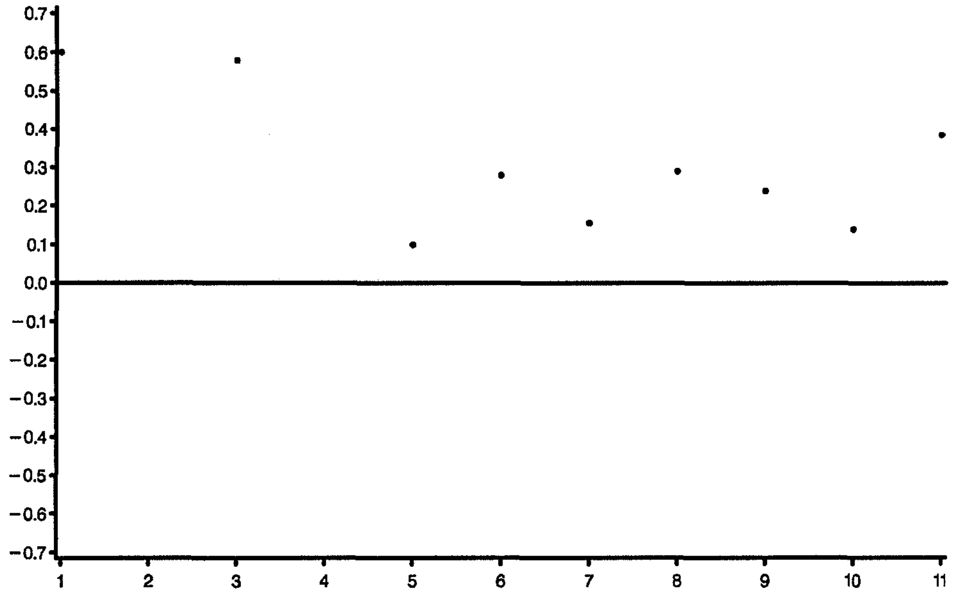


Figure 11

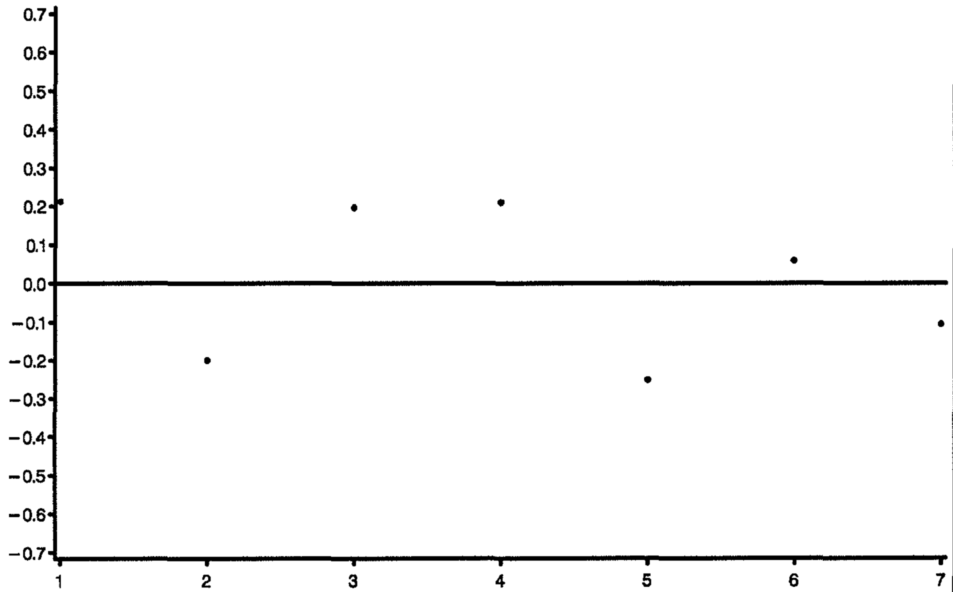


Figure 12

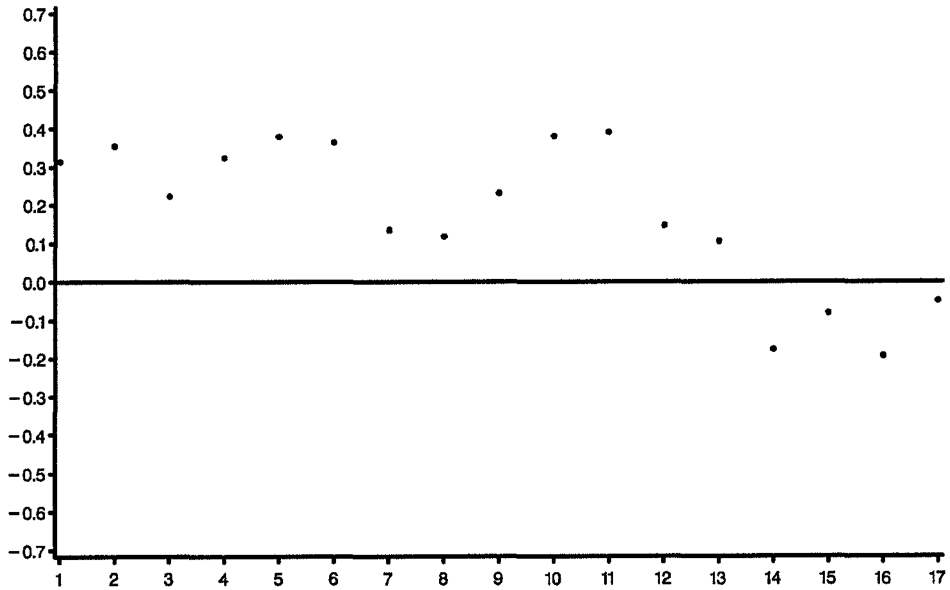


Figure 13

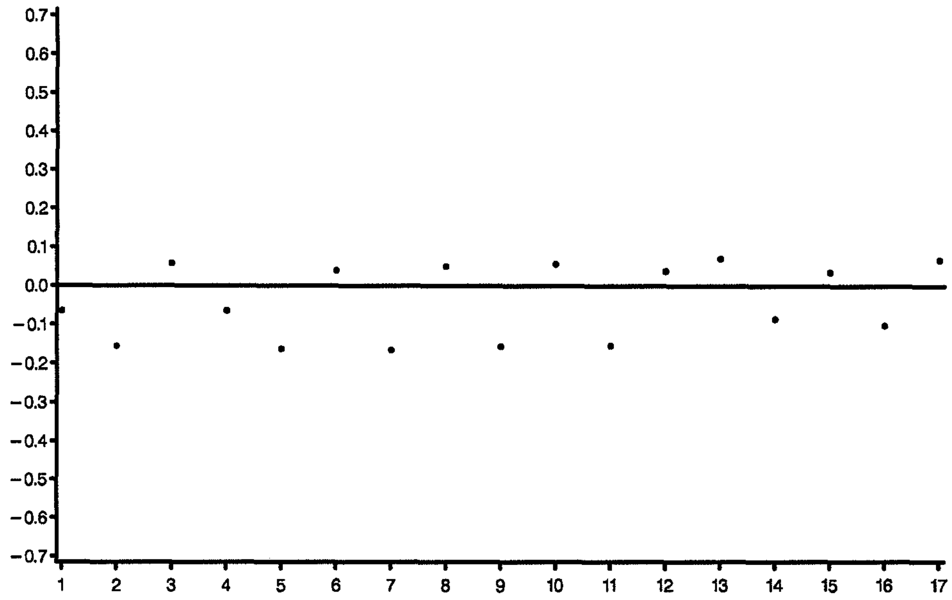


Figure 14

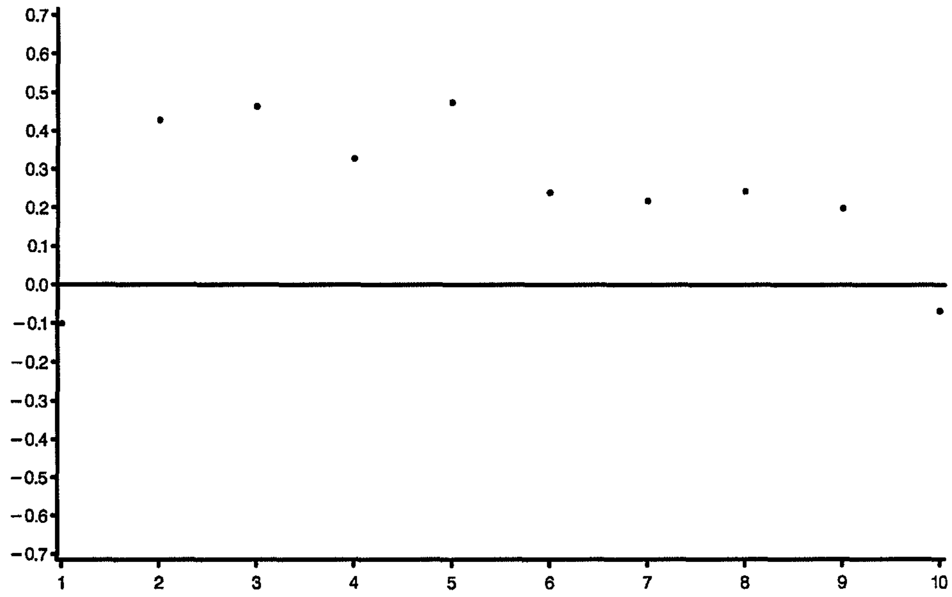


Figure 15

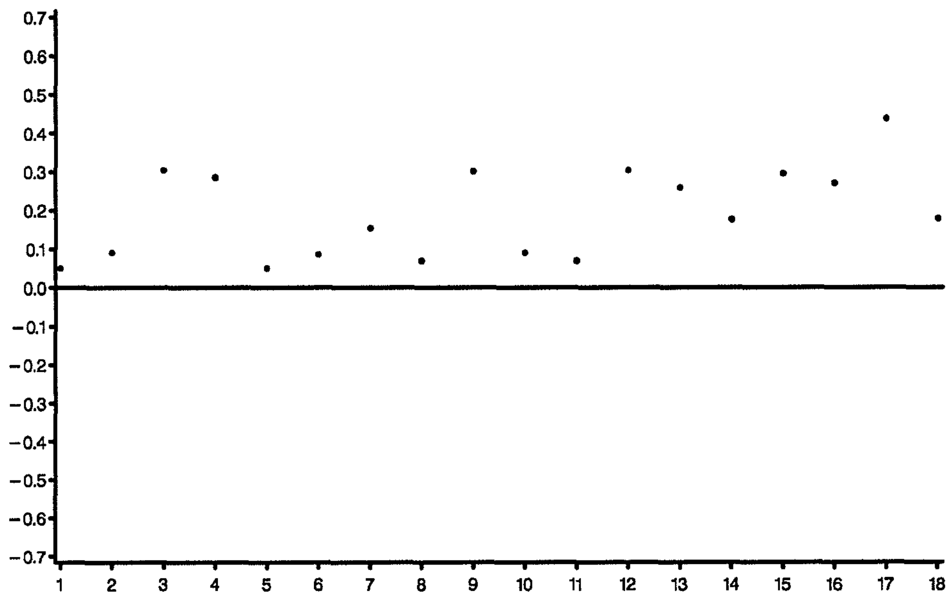
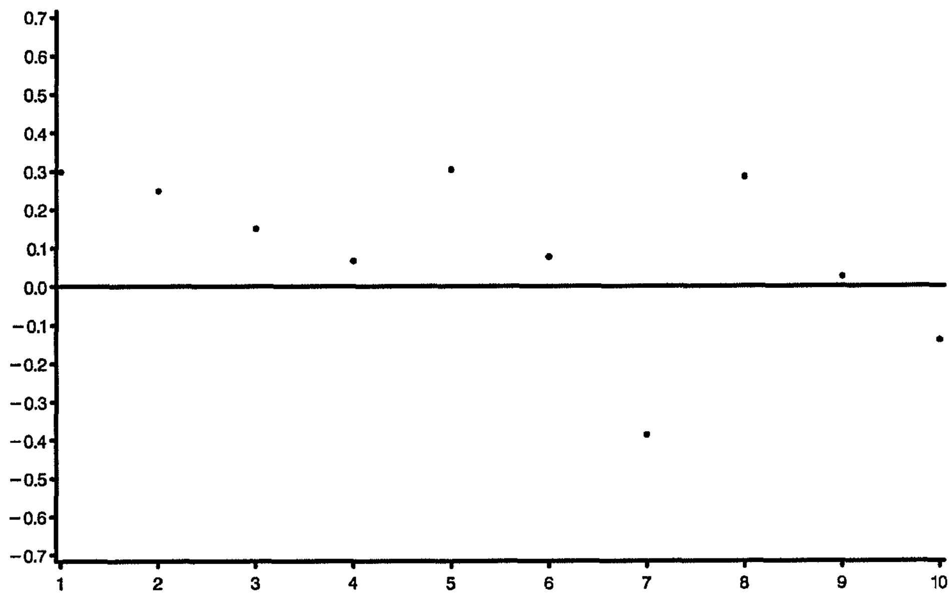


Figure 16



Research Report

2001:1	Holgersson, H.E.T.:	On assessing multivariate normality.
2001:2	Sonesson, C. & Bock, D.:	Statistical issues in public health monitoring – A review and discussion.
2001:3	Andersson, E.:	Turning point detection using non-parametric statistical surveillance. Evaluation of some influential factors.
2001:4	Andersson, E. & Bock, D.:	On seasonal filters and monotonicity.
2001:5	Andersson, E., Bock, D. & Frisén, M.:	Likelihood based methods for detection of turning points in business cycles. A comparative study.
2001:6	Sonesson, C.:	Evaluations of some exponentially weighted moving average methods.
2001:7	Sonesson, C.:	Statistical surveillance. Exponentially weighted moving average methods and public health monitoring.
2002:1	Frisén, M. & Sonesson, C.:	Optimal surveillance based on exponentially weighted moving averages.
2002:2	Frisén, M.:	Statistical surveillance. Optimality and methods.
2002:3	Jonsson, R. & Persson, A.:	Bayes prediction of binary outcomes based on correlated discrete predictors.