W.N. van Nooten

D. Draaisma

B.G.M. Ris

A MODEL FOR INDIVIDUAL SICKNESS ABSENCE FREQUENCIES

Netherlands Institute for Preventive Health Care - TNO

Leyden

August 1980

Contents

1. INTRODUCTION

A lot of research on the distribution of sickness frequencies over a group of individuals is based on the assumption that the individual sickness frequency follows a Poisson distribution. The parameter of this distribution, generally referred to by the term 'liability', is considered to be made up of two components, one determined by environment and situation factors and termed 'exposition', and one determined by internal factors inherent in the individual himself, the sickness tendency or disposition (De Cock & Corthouts, 1974). The latter component is deemed to be constant over relatively short periods of time, although usually some change is assumed to take place in the course of time. If this disposition differs from one person to another, we speak of 'absence-proneness'. With a view to investigating the presence of absence-proneness two different approaches were followed, namely the univariate and the bivariate approach. In the univariate approach it is assumed that the dispositions in the population exhibit a gamma distribution (see, for example, Froggatt,1970, hypothesis B). This leads to a distribution of sickness frequencies which is negative binomial.
Thus, in the literature on sickness absenteeism one repeatedly encounters research in which the negative binomial distribution is fitted to the distribution of sickness frequencies of groups of workers. (See, for example, Hinkle et al., 1956; Ferguson, 1972). As proof of absence-proneness this meets with the necessary objections. On the one hand a negative binomial distribution can occur as a result of assumptions completely different from those mentioned above (Shaw & Sichel, 1971, Section 11), while on the other the existence of absence-proneness, which essentially implies little more than that difference in disposition exists within a group, can lead to distributions totally different from the negative binomial distribution (viz. if the gamma distribution inadequately describes the distribution of the dispositions). What this means is that the negative binomial distribution is neither necessary nor sufficient as proof of absence-proneness.
The bivariate approach fits in with the dispositions being constant over relatively short periods of time. The inequality of the dispositions for different persons must find expression in the correlations of sickness frequencies in respect of groups of persons in two or more non-overlapping periods of time. These correlations must then be positive. Positive correlations are indeed found on numerous occasions (e.g. Hinkle et al., 1961).

It is evident from the foregoing that in order to demonstrate absence-proneness one must have certainty that the conditions are the same for all persons in the population. A guarantee of this kind is seldom given. Hence, in this article as well it is impossible to decide in favour of absence-proneness. The actual subject is, therefore, a different one: In how far is the assumption that the individual sickness frequently follows a Poisson distribution tenable?
A remarkable point is that in the literature practically no attention is given to testing this fundamental assumption.

For the present investigation we have at our disposal the sickness absence data of 3726 male manual workers, divided over five concerns, where all of them were employed during the period of eleven years between 1-1-1958 and 1-1-1969. For the purpose of the calculations the five concerns are kept separate. The distribution of the workers over the concerns is as follows:

concern	number of workers	kind of production process
1	323	processing natural oil products
2	681	cable works
3	1351	shipbuilding
4	690	shipbuilding
5	681	municipal gas and electricity company
total:	3726	

The following table presents, per concern, the total number of sickness absence spells per man-year, the number of 'never sick', i.e. those who have not been sick once throughout the entire period of eleven years, and the number of sickness absence spells of those who have been absent the most frequently (see next page).
Striking figures are the large number of 'never sick' and the low average number of sickness absence spells in concern No. 2 and generally the appreciable differences between the average number of sick reports per man-year between the concerns. This, therefore, is a very good reason for treating the concerns separately.

concern	total sickness frequency	number of sickness absence spells per man-year	never sick	maximum number of sickness absence spells
1	3113	.8776	6	40
2	5010	.6688	78	45
3	14834	.9982	15	80
4	12108	1.5953	10	105
5	9126	1.2183	13	64
total:	44991	1.0977	122	

The article is set out as follows.
Section 2 contains a proposal for a model defining the individual sickness frequency. The choice has fallen on a Poisson model, while some - otherwise familiar - properties of Poisson distribution are summed up in Section 3. It appears that allowance must be made in the model for differences between the years (Section 4) and for differences between individuals (Section 5). The latter is looked at from different angles. Ultimately, a model is proposed, in which apart from the differences described in Sections 4 and 5 no further factors are included. The significance of this is explained in Section 6 and the resulting model is tested in Section 7. But there is yet another way to test the model. This method is derived in the Appendix and the consequences for the model are discussed in Section 8. In Section 9 findings from the literature are cited for the sake of comparison, while Section 10 presents a summary.

2. THE INDIVIDUAL SICKNESS FREQUENCY

In order to draw up a model describing the individual sickness frequency we start off by postulating a probability of reporting sick in a given short period, e.g. on a given day. It may differ from day to day and from worker to worker. This probability is small and even in the case of a worker who is sick over a hundred times in the eleven years the probability is in the order of magnitude of only 0.03.

The first important assumption is that this probability is independent of the previous occurrence of sickness absence spells. Strictly speaking, this cannot be correct, since if a spell of sickness occurs, this will be of a certain duration and consequently the probability after a spell of sickness has occurred is reduced to zero for some time. On the other hand the occurrence of sickness can in some situations increase the probability of recurrence, especially in the case of certain types of sickness of a recurring nature. We have assumed, however, that such effects are slight.

Thus, the sickness frequency of a worker in a given period (e.g. one year) is defined as the cumulative effect of a large number of improbable events, which according to the above-mentioned assumption are mutually independent.

In this situation it is true to say that the individual sickness frequency over such a period approximately exhibits a Poisson distribution (Feller, pp. 234, 235; Index of Health, p. 2). A summary of some simple properties of the Poisson distributions is presented in the next Section. For the time being it is important to know that each Poisson distribution contains one unknown quantity - the parameter - which fully determines the form of the distribution.

This parameter is moreover the mean of the probability distribution, also referred to as the expectation. If for the period of one year a person has the parameter 5, it is possible to calculate for this person the probability of 0, 1, 2, ... sickness absence spells in that year. The most probable values are those about 5, e.g. 3, 4, 5, 6, 7 sickness absence spells. It is now clear that the parameter will invariably be positive, because negative numbers of sickness absence spells have no significance and we assume that anybody can be sick, thus eliminating zero expectation.

If we divide the period of eleven years into separate years, the parameter can also be expressed thus λw_i for the i-th year, where λ and w_i

are both positive. The significance of λ as a person-related value is discussed later. w_i relates to the i-th year. We shall see further on that the value of w_i is not the same for all the years. For the time being the model is still fully of a general nature, because the value of w_i has not been assumed to be the same for different workers. Hence, there is still nothing to test with reference to the model. Per period of time exactly one observation is available, viz. the sickness frequency in that period for the worker observed. At this stage nothing can be said about the mutual relations these sickness frequencies must fulfil. In order to establish what model is appropriate for the observations it will be necessary to make further assumptions in respect of the structure of w and the nature of λ.

3. SOME PROPERTIES OF POISSON DISTRIBUTIONS

A Poisson distribution is a probability distribution on the numbers 0, 1, 2, ... This probability distribution is fully determined, i.e. the probabilities of the occurrence of 0, 1, 2, ... events are determined as soon as one positive magnitude, the parameter, is fixed. If for the time being we give this parameter the notation μ, the probability of k events is as follows:

$$P(k;\mu) = e^{-\mu}\frac{\mu^k}{k!}, \ \mu > 0 \quad \text{for } k = 0, 1, 2, \ldots ,^*$$

and thus the probability value is fixed given μ for any k irrespective of the value of μ.

For a Poisson distribution the expectation, i.e. the mean of the probability distribution, is precisely μ.

Moreover, the variance of the probability distribution is likewise μ.

A fundamental property of Poisson distribution is formulated in what is known as the law of stability, which reads:

The sum of a number of independent Poisson-distributed variables itself follows a Poisson distribution with as parameter the sum of the parameters of the distributions of the constituent variables. This law enables a number of important conclusions to be drawn in respect of Poisson distributions by applying simple operations to the parameters.

For example, assuming that the sickness frequencies of a person in the eleven years are $x_1, x_2, \ldots , x_{11}$ and that these frequencies are independent and Poisson-distributed with parameters $\lambda w_1, \lambda w_2, \ldots , \lambda w_{11}$, their sum is then likewise Poisson-distributed having as parameter $\lambda(w_1 + w_2 + \ldots + w_{11})$, provided at least λ does not change in the course of the years. Hence, in this case summation takes place over the eleven part periods (years) of the total period.

---

* Here e represents the base of the natural logarithm.

4. DIFFERENCES IN PARAMETER BETWEEN THE PERIODS OF TIME

To start with we shall examine a model based on the assumption that for periods of fixed length the parameter is the same for each individual. Again we take these periods to be years, although naturally other lengths of time could also have been chosen, such as periods of six months. Taking years has the advantage that any differences found cannot result from the occurrence of seasonal influences of a systematic nature. It is a familiar phenomenon that sickness absence is more frequent in the first half of the year than in the second half.

Nevertheless, it is unlikely that the model will provide a good fit, since years differ as regards the occurrence or absence of epidemics, while moreover there is a noticeable rising trend in the number of sickness absence spells over the years under examination.

We now postulate as model that the parameter for each worker remains unchanged from year to year. According to the law of stability the sum of the sickness frequencies for all the workers then once more exhibits a Poisson distribution with the same parameter from year to year.

We now test the null-hypothesis of equality of these parameters, taking as alternative hypothesis that differences exist between the parameters, using the χ^2-test, also referred to as the dispersion or variance test (Plackett, 1974).

Table 1 presents a survey of the total sickness frequency of the five concerns from year to year, followed by the test statistic, which under the null-hypothesis has a chi-square distribution with 10 degrees of freedom.

Table 1. Total numbers of sickness absence spells for the five concerns and for the eleven years

concern	year											χ^2_{10}
	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	
1	266	325	288	246	263	266	275	286	331	255	317	29.81
2	315	417	408	329	393	407	443	425	453	562	858	483.43
3	1172	1323	1326	1193	1378	1488	1163	1356	1466	1338	1631	152.02
4	946	1197	1108	1081	1139	1226	1020	1183	1128	1098	982	71.72
5	830	917	908	800	816	854	792	842	896	641	830	68.69

All the chi-square values found are highly significant ($P < .001$), which means that important differences occur in the parameter between the years. This is the case in particular with concern no. 2 owing to a sudden rise in the number of sickness absence spells in 1967 and 1968 and to a lesser extent with concern no. 3, which exhibits a substantial rise in 1968.

From this it is clear that in the model allowance will have to be made for differences in the parameters from year to year.

5. DIFFERENCES IN PARAMETERS BETWEEN THE PERSONS

We have seen that the parameter depends on the period examined. The question that now also arises is whether differences in parameter also occur from person to person. Such differences are indeed likely, because if they did not occur and thus the mean of the probability distribution of sickness frequencies would be the same for all persons, there would also be no variables bearing a relation to the sickness frequencies. But such variables definitely exist.

One variable of this kind is the year of birth. Further variables encountered by Philipsen (1977) were whether a person has had an eventful life, alienation, need for leisure time, smoker or non-smoker.

A second indication is found in the existence of positive correlations of the sickness frequencies between the different years. These would, of course, also have to be zero, if there were no differences in parameter from person to person. At the end of this Section we shall show a correlation matrix for our material.

In analogy with our reasoning when comparing the annual data we shall now perform a test to examine whether the null-hypothesis to the effect that all persons have the same parameter is tenable. The alternative hypothesis states that differences in parameter occur. The test applied is the same as that referred to in the previous Section, except that the number of degrees of freedom of the calculated χ^2-value is so great as to enable it to be transformed into a standard-normal test statistic T as follows: $T = \sqrt{2\chi^2} - \sqrt{2df - 3}$, where df is the number of degrees of freedom. The test was performed for the five concerns separately for the year 1960.

Table 2 shows the χ^2-values, the numbers of degrees of freedom df, the standard-normal test statistic T and the level of significance, under P.

Table 2. Test results for the null-hypothesis of one and the same parameter for all individuals in respect of the five concerns in 1960

	concern				
	1	2	3	4	5
χ^2	458.94	882.23	1842.63	1051.68	1033.00
df	322	680	1350	689	680
T	4.939	5.141	8.755	8.754	8.589
P	<<.01	<<.01	<<.01	<<.01	<<.01

In each of the cases the test result is that the null-hypothesis must be rejected. Hence, there are indeed differences in parameter between the individuals.

We shall now illustrate this effect in a different way. Again let us assume that the sickness frequency of all persons is based on a probability distribution with the same expectation and that these probability distributions are Poisson distributions. The common expectation fully determines the probability distributions and consequently all the persons now have an identical probability distribution. The sickness frequencies then constitute a sample from a single Poisson distribution. Table 3 shows the distributions of numbers of sickness absence spells for the year 1960 in respect of the five concerns. In addition, the appertaining best-fitting Poisson distribution is also represented. The deviations appear to be considerable. The distribution observed invariably exhibits a greater variance than the adapted Poisson distribution, since for the latter the mean and the variance must after all be equal. This is definitely not the case here.

Table 3. Frequency distributions of number of sickness absence spells (a) and best-fitting Poisson distribution (b) for the five concerns in 1960

	concern									
frequency	1		2		3		4		5	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
0	151	132.42	449	374.07	576	506.29	199	138.50	236	179.51
1	103	118.07	132	224.11	429	496.92	183	222.41	198	239.35
2	39	52.64	55	67.14	211	243.86	147	178.57	122	159.56
3	19	15.65	25	13.41	94	79.78	77	95.58	75	70.92
4	7	3.49	10	2.01	27	19.58	46	38.37	24	23.64
5	3	.62	9	.24	8	3.84	21	12.32	18	6.30
6	-	.09	1	.02	3	.63	12	3.30	5	1.40
7	1	.01	-	.00	1	.09	3	.76	1	.27
8	-		-		-	.01	1	.15	1	.04
9	-		-		1	.00	-	.03	-	.01
10	-		-		-		1	.00	1	.00
11	-		-		1		-		-	
total	323		681		1351		690		681	
mean	.89		.60		.98		1.61		1.33	
variance	1.27		1.11		1.34		2.45		2.03	

When examining the distributions, it appears that compared with the adapted distribution (b) the distribution observed (a) invariably exhibits too many zeros and too many high values. It is precisely this state of affairs which is responsible for the excessively large variance in (a).

It can be demonstrated that the variance is made up of two components. One component is due to the Poisson properties of the individual sickness frequency and the other to the differences in parameter between the individuals. This explains why the variance of the observed distribution must be *greater* than the mean, which is an estimate of the first component alone.

The correlation matrices of the sickness frequencies between the eleven years were then calculated for the five concerns examined. The result for concern no. 3 is shown in table 4.

Table 4. Correlation matrix of the sickness frequencies between the eleven years in respect of concern no. 3 (n = 1351)

year	year										
	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968
1958	1										
1959	.43	1									
1960	.40	.38	1								
1961	.37	.41	.44	1							
1962	.40	.38	.45	.45	1						
1963	.40	.40	.39	.42	.46	1					
1964	.28	.34	.34	.35	.37	.47	1				
1965	.30	.32	.31	.34	.39	.41	.43	1			
1966	.36	.36	.37	.39	.40	.46	.44	.47	1		
1967	.27	.30	.29	.32	.38	.39	.39	.44	.48	1	
1968	.26	.26	.27	.28	.35	.36	.32	.40	.43	.40	1

The reason for including this correlation matrix in the discussion is as follows. We know that in a given period differences in parameter exist between individuals. These differences will be due to the occurrence of conditions and circumstances exhibiting a relation with the sickness frequency and of which it would be true to say, therefore, that they contribute to the parameter of the probability distribution.

If these conditions and circumstances are permanent, the parameter does not change, at least as far as the individual component is concerned. Hence, the situation is as follows. As a result of individual differences in conditions and circumstances a dispersion of parameter values occurs, while insofar as these conditions and circumstances are permanent the relations between the parameters remain the same. Consequently, sickness frequencies for individuals over different periods will exhibit a positive relationship.

The results of concern no. 3 are characteristic of all five concerns. It is generally true to say that all the correlations found are positive, which is in agreement with the assumption of individual differences in parameter values of a semi-permanent character. Furthermore, it appears that the height of the correlations gradually decreases according as years spaced farther apart are compared. The person parameters are apparently not strictly constant and subject to slight shifts in the course of time. There are also some characteristic differences between concerns. In the case of concern no. 1 the correlations are generally lower and exhibit greater fluctuations than those of concern no. 3. Concern no. 2 shows somewhat higher correlations, but a sudden drop occurs in 1967 and 1968, thus again indicating that something must have happened in this concern after 1966. Concerns nos. 4 and 5 exhibit the same picture as concern no. 3, except that both concerns, but no. 5 in particular, exhibit somewhat lower relationships.

6. SPLITTING OF THE PARAMETER

The foregoing leads to the conclusion that at least two sources of variation must be taken into account. On the one hand there are differences between the years and on the other between individuals, which must be ascribed to the parameters of the distributions controlling the sickness frequencies. A next step is, therefore, to notate these parameters in a form that makes allowance for both the year effects and the individual differences. This requirement is fulfilled, if for person r in period i we notate the parameter thus: $\lambda_r w_i$, when each person contributes his own value λ_r and each period its value w_i to the parameter. We shall now ascertain the value of the model so obtained. The model is no longer as general as possible. It lacks a factor which is a mixture of an individual and a period contribution. Such a factor would have to be introduced in the event of interactions taking place between individuals or groups of individuals and periods. Such interactions can occur if individuals react differently to changes in the course of time, which are the same for all or if changes occur which apply to some workers and not to others. In view of this the notation amounts to the introduction of an important assumption, of which the tenability is examined by means of the following test. But first we shall discuss its significance.
To start with it should be pointed out that the introduction of a factor relating to a combination of individuals and period does not impose limitations on the model.
Any distribution of individual sickness frequencies found over the periods can be explained by an appropriate choice of this factor. For the sake of discussion, however, we now introduce this form for the parameter: for individual r in the i-period the parameter may be notated thus: $\mu_{ri} = \lambda_r d_{ri} w_i$.
λ_r now brings to expression that a collection of influences exists, which is of a permanent character and coupled to the individual. Examples are: the year of birth and, provided they remain unchanged, the family situation, the general state of health, the concern and the department within the concern and so on. In w_i all period effects are accounted for insofar as they equally apply to all individuals.
Situations that come to mind in this respect are epidemics and changes in the concern applying to all persons.
d_{ri} relates to the previously mentioned interactions. The emphasis here can be placed on either the individual aspect or the period aspect,

although the distinction is somewhat artificial. We are concerned with the individual aspect in particular when some members of a group, for which the period-related conditions change in identical manner, react differently to this change than others. Reorganization plans for instance can have a different effect on older than on younger people. The period aspect must be taken into account when changes occur in the course of time which apply to some individuals and not to others. The appointment of a new head of a department for example. Apart from this there is yet another category of influences that cannot be placed under one of these two headings. Age and years of service rise as the years go by and the changes affecting the parameter as a result of this can be of fluctuating significance according as the starting value differs. Assuming in the development of the model that $d_{ri} = 1$, this implies that we assume that the categories of interaction referred to have no real effect on the parameter. This assumption can be met by examining the model in relation to groups of individuals for whom the conditions are 'epidemiologically homogeneous' (Nass, 1956). This means to say that the period aspect of the interaction is as far as possible eliminated by an appropriate choice of group examined. The other aspects of the interaction cannot be controlled. We have no means of applying a selection capable of guaranteeing that the selected individuals will react to changes in an identical manner. Starting age and year of joining could be made constant by selection, but in our material this would lead to a substantial reduction of the number of cases examined. Moreover, no information whatever is available concerning years of service. It should also be pointed out that an important variable, related to (but not identical with) age, viz. the year of birth, is automatically kept constant by the nature of the data. After all, each individual is compared with himself in the course of time and, of course, his year of birth remains the same throughout the entire period. We hope that the interactions, although perhaps not entirely absent, will prove to be of such minor importance as to eliminate noticeable disturbance of our model.

7. TESTING THE MODEL

The model, of which we are going to examine the application, can now ultimately be defined as follows:
the sickness frequency of individuals r in period i follows a Poisson distribution, of which the parameter can be notated as $\lambda_r w_i$ *for any r and* $i = 1, 2, \ldots, n$. *These distributions are mutually independent from period to period and from individual to individual.*
Hence, it is assumed here that the interaction plays no part at all. In this situation use can be made of the chi-square test for a kxn table, proviced the expectations in the kxn cells are not too small*.
If we were to consider all the individuals by themselves, this latter condition would definitely not be fulfilled. To solve this problem the individuals were, where necessary, gathered together in groups and their joint sickness frequency examined.
To this end the groups were composed as follows: all persons with in all 1, 2, 3, ... sickness absence spells were taken together. If it occurred that the total sickness frequency of such a group over eleven years was then still lower than about 60, some of these groups with different totals were again combined. Persons who by themselves had 60 or more sickness absence spells were invariably kept separate (groups of one person). On the strength of the law of stability the sickness frequency for such a group in each period i again displays a Poisson distribution with as parameter the sum of $\lambda_r w_i$, taken over the r's belonging to the group. The result is that the parameter takes the form of $\lambda_g w_i$, where λ_g is the sum of the individual λ_r appertaining to the group. Hence, gathering into groups does not upset the model.
The test is performed per concern, thus five times. Table 5 gives a survey of the χ^2-values found, the numbers of degrees of freedom df, the approximately standard-normal test statistic $T = \sqrt{2\chi^2} - \sqrt{2df - 3}$ and the appertaining tail probability P.

---

* An effort has been made to group the data so that the expectation per cell is nowhere much lower than 5.

Table 5. Results of testing the model on five concerns

concern	chi-square	df	T	P
1	273.57	250	1.098	.14
2	934.38	340	17.210	$<10^{-10}$
3	432.43	450	-.541	.70
4	552.55	530	.731	.23
5	418.54	410	.001	.50

With four out of the five concerns the model proves to be well in agreement with the data. Concern no. 2 is an exception in this respect. The chi-square value found indicates that the model is absolutely incompatible with the data.

We may now try to ascertain what is wrong with concern no. 2. Previously we found for this concern a substantial increase in the number of sickness absence spells in the last two years of the period examined and the correlation matrix of the sickness frequencies between the years revealed a level drop in the correlations between the first nine and the last two years compared with those within these two groups of years.

The frequencies found for concern no. 2 and grouped in the manner described were entered in table 6 and supplemented with the appertaining expected value of the chi-square test.

The contribution per cell to the total chi-square value is shown between brackets. In connection with the size of the table only a part is shown. This table is also an illustration of the operations carried out in respect of all the concerns.

Considering that the situation in the part of the table not shown is roughly as represented in the last part, it will be clear that the cause of the incompatibility of the model for the last two years must be sought in the contradiction between the group with a small number of cases of absenteeism (up to about 6 in eleven years) and that with more cases.

Table 6. Sum of the number of sickness absence spells of groups of individuals with in all k absence spells over the years 1958 - 1968, expected values with relevant model and contributions to chi-square. Concern no. 2

k	'58	'59	'60	'61	'62	'63	'64	'65	'66	'67	'68	total
	0	0	2	0	1	1	2	1	3	22	45	77
1	4.84	6.41	6.27	5.06	6.04	6.26	6.81	6.53	6.96	8.64	13.19	
	(4.84)	(6.41)	(2.91)	(5.06)	(4.21)	(4.42)	(3.40)	(4.69)	(2.25)	(20.67)	(76.75)	(135.59)
	2	4	4	3	6	2	5	4	9	30	91	160
2	10.06	13.32	13.03	10.51	12.55	13.00	14.15	13.57	14.47	17.95	27.40	
	(6.46)	(6.52)	(6.26)	(5.36)	(3.42)	(9.31)	(5.91)	(6.75)	(2.07)	(8.09)	(147.61)	(207.76)
	9	5	5	4	7	10	8	8	21	29	50	156
3	9.81	12.98	12.70	10.24	12.24	12.67	13.79	13.23	14.11	17.50	26.72	
	(.07)	(4.91)	(4.67)	(3.81)	(2.24)	(.56)	(2.43)	(2.07)	(3.37)	(7.56)	(20.29)	(51.98)
	9	9	7	8	9	9	10	9	14	27	73	184
4	11.57	15.31	14.98	12.08	14.43	14.95	16.27	15.61	16.64	20.64	31.51	
	(.57)	(2.60)	(4.25)	(1.38)	(2.05)	(2.37)	(2.42)	(2.80)	(.42)	(1.96)	(54.62)	(75.44)
	9	14	15	5	11	14	17	15	19	35	56	210
5	13.20	17.48	17.10	13.79	16.47	17.06	18.57	17.81	18.99	23.56	35.96	
	(1.34)	(.69)	(.26)	(5.60)	(1.82)	(.55)	(.13)	(.44)	(.00)	(5.56)	(11.16)	(27.56)
	10	21	10	9	16	11	6	10	17	29	53	192
6	12.07	15.98	15.64	12.61	15.06	15.60	16.98	16.29	17.36	21.54	32.88	
	(.36)	(1.58)	(2.03)	(1.03)	(.06)	(1.36)	(7.10)	(2.43)	(.01)	(2.59)	(12.31)	(30.84)
	10	22	17	9	10	11	12	16	18	25	39	189
7	11.88	15.73	15.39	12.41	14.83	15.35	16.71	16.03	17.09	21.20	32.37	
	(.30)	(2.50)	(.17)	(.94)	(1.57)	(1.23)	(1.33)	(.00)	(.05)	(.68)	(1.36)	(10.12)
	4	8	11	7	7	5	8	8	3	5	4	70
35	4.40	5.83	5.70	4.60	5.49	5.69	6.19	5.94	6.33	7.85	11.99	
	(.04)	(.81)	(4.93)	(1.26)	(.41)	(.08)	(.53)	(.72)	(1.75)	(1.04)	(5.32)	(16.88)
36	7	11	5	7	6	5	9	6	4	5	7	72
	4.53	5.99	5.86	4.73	5.65	5.85	6.37	6.11	6.51	8.08	12.33	
	(1.35)	(4.18)	(.13)	(1.09)	(.02)	(.12)	(1.09)	(.00)	(.97)	(1.17)	(2.30)	(12.43)
	10	12	9	10	12	11	8	9	8	15	13	117
39	7.36	9.74	9.53	7.68	9.18	9.50	10.35	9.93	10.58	13.12	20.04	
	(.95)	(.53)	(.03)	(.70)	(.87)	(.24)	(.53)	(.09)	(.63)	(.27)	(2.47)	(7.29)
	4	3	4	5	7	10	8	9	6	13	11	80
40	5.03	6.66	6.51	5.25	6.28	6.50	7.07	6.79	7.23	8.97	13.70	
	(.21)	(2.01)	(.97)	(.01)	(.08)	(1.89)	(.12)	(.72)	(.21)	(1.81)	(.53)	(8.57)
	4	4	10	5	7	9	9	11	11	10	6	86
41,45	5.41	7.16	7.00	5.65	6.75	6.99	7.60	7.30	7.78	9.65	14.73	
	(.37)	(1.39)	(1.28)	(.07)	(.01)	(.58)	(.26)	(1.88)	(1.34)	(.01)	(5.17)	(12.36)
total	315	417	408	329	393	407	443	425	453	562	858	5010

In terms of interactions: a change must have occurred in the years 1967/1968, to which 'low-absence' workers reacted differently from 'high-absence' workers in that the former category displayed a much greater rise in the number of people reporting sick. In concern no. 2 there is apparently interaction between the periods and the total sickness frequency per individual.

It was ascertained whether this interaction could be ascribed to the age composition and whether there were any differences in the duration distributions of the sickness absence spells. No further explanatory variables were available.

In respect of all workers in concern no. 2 with 10 or fewer sickness absence spells these absence spells were divided into those which occurred in the years 1958/1966 and those which occurred in 1967/1968. It was then ascertained in respect of all these cases what the age cohort was of those responsible for them. This produces table 7 of frequencies appertaining to the age cohorts and the year groups.

Table 7. Division according to age cohort and year groups of sickness absence spells appertaining to workers in concern no. 2 with at most 10 absence spells in all

age cohort	year groups		total
	1958/1966	1967/1968	
1	49	27	76
2	60	58	118
3	137	105	242
4	191	167	358
5	254	196	450
6	223	143	356
7	143	99	242
total	1057	795	1852

This table has a chi-square value of 7.189 at 6 degrees of freedom and a tail probability P = .30. Hence, there is no indication of a relationship existing between age cohort and the year group in which the absence spells occurred.

For the same category of workers in concern no. 2, with a total sickness frequency of at most 10 in the eleven years, a comparison was made between the duration distribution of their absence spells in the two year groups. The absence spells were subdivided into those of at most a week and those of longer duration. This resulted in a table of frequencies - table 8 - arranged according to duration and year groups.

Table 8. Distribution according to duration and year group of sickness absence spells in respect of those workers in concern no. 2 with a total of at most 10 absence spells

duration	year groups		total
	1958/1966	1967/1968	
up to 1 week	285	259	544
longer than 1 week	772	536	1308
total	1057	795	1852

This table has a chi-square value of 6.897 at 1 degree of freedom and a tail probability $P < .01$. When examining the table it is evident that a relatively greater number of sickness absence spells of short duration occurred in 1967/1968. This exhausts our possibilities for elucidation of the interaction effect found. The figures for workers with a low total number of absence spells show a substantial rise in the last two years considered. The age composition has nothing to do with this, but there is a certain tendency towards the occurrence of short absence spells in the last two years. However, this certainly does not exhaustively explain the interaction effect.

An important point is, however, that the model can serve as an aid to demonstrate the presence of and locate such interactions.

8. CONSEQUENCE OF THE MODEL

It was evident from the foregoing that the model led to consequences capable of being tested and that with four out of the five concerns the test results did not give any reason to reject the model. In the case of concern no. 2 we found substantial deviations and an indication was given as to where these deviations occurred. Hence, the model acquired the function of an instrument to detect deviations in the sense of interactions and make them stand out in order that they might be explained.

In this way we endeavoured to ascertain whether verification of the model had already taken place to an adequate extent, but in actual fact this was not yet the case. Our findings in respect of concern no. 2, however, do not constitute sufficient reason to reject the model. What they do in fact imply on the other hand is that the model must be applied with due care if changes of a drastic nature occur in the course of the years. It would also appear to be important regularly to present findings obtained with the model in respect of new material, a point made by Nass as far back as 1956, but which was not followed up.

The function of the model will now once more be highlighted in an application demonstrating that it cannot be correct in detail, but at the same time proving its value as a 'touchstone'.

For one of the concerns the correlations of the sickness frequencies, calculated in the usual way and based on the simultaneous distributions of the sickness frequencies for pairs of years, are given in Section 5. We shall call these correlations the 'experimental correlations'. In addition, the model offers the possibility of calculating 'theoretical correlations', enabling these correlations between pairs of years to be estimated on the strength of the frequency distributions of these years without making use of the simultaneous distribution. This possibility is based on some properties of the model, viz.:

a) the individual sickness frequencies follow a Poisson distribution;

b) these sickness frequencies are independent from year to year;

c) the person parameter λ is the same from year to year.

It is demonstrated in the Appendix that when these properties are effective the correlations assume a special form, which can be estimated according to the formula:

$$r(x_i,x_j) = \sqrt{(1-\frac{\bar{x}_i}{s^2(x_i)}) \times (1-\frac{\bar{x}_j}{s^2(x_j)})},$$

where $r\ (x_i,\ x_j)$ stands for the theoretical correlation of the sickness frequencies between the years i and j, $s^2(x_i)$ and $s^2(x_j)$ being the variances calculated in the distribution of the frequencies in the periods i and j and $\bar{x}_i$ and $\bar{x}_j$ the appertaining means.

It will be seen that $r(x_i,\ x_j)$ is the product of two factors, one of which is fully determined by the distribution of the sickness frequencies in year i and the other by those in year j. The correlation will be zero if $\bar{x}_i = s^2(x_i)$ or if $\bar{x}_j = s^2(x_j)$. In that case the variance of the sickness frequencies is, at least in one of the years, equal to their mean, and this implies that the distribution fulfils the properties of a single Poisson distribution, when the parameter would thus be the same for all the workers, since the mean and the variance are then equal. The correlation becomes greater according as the ratio between the mean $\bar{x}_i$ or $\bar{x}_j$ and the variance $s^2(x_i)$ or $s^2(x_j)$ is smaller. In effect, $\bar{x}_i$ and $\bar{x}_j$ indicate what the variance would be if everybody had the same parameter, while s_i^2 and s_j^2 indicate the actual variance, which is greater according as the parameters of the individuals display greater differences.

In table 9 the experimental and theoretical correlation coefficients, again for concern no. 3, are presented alongside each other. We find that the experimental correlations for pairs of years with a small interval in between are greater than the appertaining theoretical coefficient.

If the interval is large, the theoretical coefficients are greater than the experimental ones. Both effects occur with all the concerns.

We shall now give an explanation for these two effects.

As previously pointed out in Section 5 the reduction of the correlations according as the time interval is greater can be due to shifts in the individual parameter in the longer term. This, therefore, is not strictly permanent, but semi-permanent. It is understandable that an effect of this kind should occur, since absence-influencing conditions will change in the course of time, affecting first one worker and then another.

Table 9. Correlation matrix with experimental and theoretical coefficients of the sickness frequencies between the eleven years in respect of concern no. 3 (n = 1351). The theoretical value is placed between brackets

year	year										
	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968
1958	1										
1959	.43	1									
	(.34)										
1960	.40	.38	1								
	(.32)	(.29)									
1961	.37	.41	.44	1							
	(.35)	(.31)	(.29)								
1962	.40	.38	.45	.45	1						
	(.38)	(.34)	(.32)	(.35)							
1963	.40	.40	.39	.42	.46	1					
	(.36)	(.33)	(.30)	(.33)	(.36)						
1964	.28	.34	.34	.35	.37	.47	1				
	(.38)	(.34)	(.32)	(.35)	(.38)	(.36)					
1965	.30	.32	.31	.34	.39	.41	.43	1			
	(.37)	(.33)	(.31)	(.34)	(.37)	(.35)	(.37)				
1966	.36	.36	.37	.39	.40	.46	.44	.47	1		
	(.37)	(.33)	(.31)	(.34)	(.37)	(.35)	(.37)	(.36)			
1967	.27	.30	.29	.32	.38	.39	.39	.44	.48	1	
	(.36)	(.33)	(.30)	(.33)	(.36)	(.35)	(.36)	(.35)	(.35)		
1968	.26	.26	.27	.28	.35	.36	.32	.40	.43	.40	1
	(.35)	(.32)	(.30)	(.32)	(.35)	(.34)	(.35)	(.34)	(.34)	(.34)	

The short-term effect is more surprising, since the experimental correlation is then greater than its theoretical counterpart. Taking the theoretical coefficient as a basis, a higher experimental correlation will indicate that the sickness frequency in successive years is more constant than prescribed by the model. In terms of model-assumptions this means either that the distribution of the individual sickness frequency does not follow a Poisson distribution, but one for which the variance is smaller than the expectation, or that the frequencies from year to year are *not* independent and that changes in these frequencies are less pronounced than would be the case with independence, or both. In effect, the two explanations are related.

Both indicate a positive dependence of short-term individual sickness frequencies.

We now ask ourselves to what extent the sickness frequency can be explained with the aid of a number of other variables. Again two approaches are possible: one based on the model-assumptions and one not based on these assumptions.

The best that can be achieved is to find a number of variables which have a multiple correlation of 1 with λw_i. What this amounts to is that in effect we possess all the relevant information in respect of λw_i itself.

Hence, we want to know something about the correlation of λw_i with x_i, because this gives us the maximum of the multiple correlations of explanatory variables with x_i. In the Appendix it is concluded that the correlation of λw_i with x_i, when using only the notation λw_i for the parameter (thus in the absence of interactions), is capable of being estimated as the square root of the ordinary correlation of the sickness frequencies in the years x_{i-1} and x_i. Use is made of successive years, because in such a short term we have the best guarantee that λ has not yet undergone any drastic change.

When making use of the model-assumptions, the correlation coefficient $r(x_i,\lambda w_i)$ can, as set out in the Appendix, be estimated as:

$$r(x_i,\lambda w_i) = \sqrt{(1 - \frac{\bar{x}_i}{s^2(x_i)})}$$

We shall refer to the two estimates as the experimental and the theoretical explainability of x_i respectively.

The latter is apparently identical with one of the factors of the theoretical correlation coefficient between x_i and x_j and thus corresponding properties apply.

Hence, $r(x_i,\lambda w_i) = 0$ if $\bar{x}_i = s^2(x_i)$ and $r(x_i,\lambda w_i)$ approaches unity according as $s^2(x_i)$ increases in relation to $\bar{x}_i$.

When discussing the theoretical and the experimental correlation, it also became apparent that the latter was greater than the former for years not too far apart. This implied that the model-assumptions were not altogether valid in that there was evidence of short-term dependence. This effect will also find expression here in a lower value of the theo-

retical explainability compared with its experimental counterpart.
For each enterprise the correlation between x_{i-1} and x_i was calculated for all years i and the mean $\bar{r}$ taken from these results.
The square root of this is shown in column 1 of table 10 as indication of the order of magnitude of the explainability of the sickness frequency by other variables. Similarly, a coefficient was calculated per concern on the basis of the model. To this end the sum was taken of $\bar{x}_i$ and also that of $s^2(x_i)$ for all the years i, m and s^2 respectively. Subsequently, $\sqrt{1 - \frac{m}{s^2}}$ was calculated and included in column 2 of table 10.

Table 10. Values of the square root of $\bar{r}$ and of $1 - \frac{m}{s^2}$ as indication of the explainability of the sickness frequency

concern	$\sqrt{\bar{r}}$	$\sqrt{1 - \frac{m}{s^2}}$
1	.572	.503
2	.713	.588
3	.665	.567
4	.684	.638
5	.665	.495

The coefficients calculated on the basis of the model are indeed invariably smaller than the estimate based on the correlations, which once more demonstrates that the variance per person is smaller than the ruling per-person Poisson distribution would lead us to expect.
On the strength of these findings multiple correlations in excess of about .70 are not to be expected for these concerns.

9. COMPARISON WITH RELIABILITY RESEARCH BY OTHERS

Little is found in the literature concerning the reliability and/ or stability of individual absence frequencies, a remark recently also made by, inter alia, Chadwick-Jones et al (1971 and Muchinsky (1977). The sources known to us are listed in table 11*.

The correlation coefficients mentioned here are of the same order of magnitude as those we have found. The different authors are inclined to consider the reliability rather low on the strength of their findings. In the light of our results with the model reliabilities cannot be expected to be much higher with a test and retest period of one year, as also applied by us. Associated with the tendency to underestimate the reliability found in respect of the absence frequency, there is also a noticeable tendency to depreciate the value of an established correlation coefficient of explanatory variables with the absence frequency. Nicholson & Goodge (1976) for instance have found correlations of age with absence frequency in two successive years of -.40 and -.47 respectively. All other relations with the absence frequency they examined are lower. Our findings suggest a maximum multiple correlation coefficient of approximately .70. In view of this zero-order correlations of from .40 to .50 must be considered to be quite satisfactory.

The generalizability of our findings is not clear. It is possible after all that all this only applies by the grace of the material that happened to be at hand. All the same the limited data from table 11 leave the impression that in that material empirical correlation coefficients of the absence frequencies behave roughly in the same way from one period to another. Since the means and variances of the absence-frequency distributions are not given in the sources concerned, with the exception of Morgan & Herman (1976), only in this instance we are in the position to estimate the theoretical correlation coefficient between the years and the maximum multiple correlation coefficient.

Morgan & Herman found mean m = 13.6 and standard dev iation sd = 12.47 for the absence frequencies during the 17 months preceding their study and m = 6.48, sd = 5.81 in the 7 months following it. On an annual basis

---

* Muchinsky (1977) also mentions Turner (1960) and Latham & Pursell (1975) as sources, but in these publications the measure of absence is not the frequency, as he suggests.

the absence frequencies are about 9.6 and 11.1. These are rather high figures. Furthermore, the standard deviations look very high as well. An explanation could be that they investigated a broad range of absence behavior, including problems with transportation to the plant, family illness and job disciplinary action. Using our formula for the theoretical correlation coefficient we find $r = .86$ as compared with an experimental value of .70. The theoretical explainabilities are .96 for the 17 month period and .90 for the 7 month period. The experimental value is .84. Apparently our model does not hold here, possibly due to rapid changes in λ for some causes of absence.

Table 11. Investigation of the reliability of absence frequencies from the literature

researchers	criterion	n	reliability	periods	population
1. Huse & Taylor (1962)	test-retest	393	.61	2 successive years	male truck drivers of large oil company
2. Nicholson & Goodge (1976)	test-retest	303	.57	2 successive years	female hourly paid workers of a food processing plant
3. Waters & Roach (1971)	test-retest	62	.55	2 successive years	female clerical workers
4. Waters & Roach (1979)	test-retest	82	.62	2 successive years	female clerical employees
5. Morgan & Herman (1976)	test-retest	48	.70	17 months before 7 months after the study	employees of an automobile parts foundry
6. Chadwick-Jones et al. (1971)	test-retest	318	.43	2 corresponding periods of 39 weeks	employees producing car components
7. Lyons (1968)	test-retest	33	.63	2 successive quarters	registered nurses
8. Farr et al. (1971)	Spearman-Brown	153	.39	3 months	female toll collectors

10. SUMMARY

When investigating sickness absenteeism it is important to have models describing the different aspects of this absenteeism. In the foregoing an attempt has been made to formulate such a model for the individual sickness frequencies. First if all it was demonstrated that in a model of this kind allowance must be made for differences between the years. Further, we found that there were also differences between individuals. In conformity with these findings a model was formulated, which implied that a sickness frequency can be regarded as a drawing from a Poisson distribuation with parameter $\lambda_r w_i$, where λ_r is characteristic of person λ_r,w_i for periode i. With four out of the five concerns examined an initial test proved the model to be well in agreement with the observations. The only deviating concern illustrated that the model can be used to detect interactions between years and individuals. With this method such interactions can be discovered and located. The differences existing between individuals are of a semi-permanent nature. Consequently, each individual has his own level of absence frequency and a certain amount of explainability exists with respect to this frequency. In terms of maximum attainable multiple correlations this explainability varies from approximately .50 for concerns nos. 1 and 5 and to approximately .64 for concern no. 4, when using the model-assumptions. However, it appears that these model-assumptions are not always strictly valid. In effect, the sickness frequencies appear to exhibit dependence in the short term, which comes to expression in excessive constancy of the sickness frequency per individual.
This results in an increased explainability of the sickness frequencies. Nevertheless, this explainability does not yet exceed at most .71 even now, corresponding at best to roughly 50% explained variance.
All in all the model has proved to function reasonably well. The deviations observed in the form of interactions in concern no. 2 are considerable, but those resulting from the dependence referred to above are concerned with details. We feel it would be a good thing if some of the operations reported here could be repeated whenever material becomes available that permits of these analyses being made. In this way the usefulness of the model and its limitations can be brought to light.

APPENDIX

With x_{ir} we denote the sickness frequency of person r in year i. x_{ir} is composed of a 'true score' t_{ir} increased by a chance fluctuation e_{ir}. For person r t_{ir} is a fixed value, e_{ir} a drawing from a probability distribution with a mean o. The Poisson model further implies that the variance $\sigma^2(e_{ir}) = t_{ir}$ and that in general the probability distribution of e_{ir} depends on t_{ir}, thus also on the person r.
We regard the formation of the frequency x_{ir} as a process that is brought about in two stages. In the first stage a value t_{ir} is taken at random from a distribution over values t_i, while in the second stage, given t_{ir}, a value e_{ir} is taken at random from the distribution appertaining to t_{ir}. After this description we can notate in a more general sense: $x_i = t_i + e_i$, where t_i is a stochastic variable with values t_{ir}, and e_i with values e_{ir} can be realized according to the stepwise process.
For the different years i = 1, 2, ... k we thus have stochastic vectors* $x = (x_1, x_2, ..., x_k)^T$ and in analogy $t = (t_1, t_2, ..., t_k)^T$ and $e = (e_1, e_2, ..., e_k)^T$. Thus: $x = t + e$.
In this situation the following identities can be used to calculate expectation E and covariance matrix Σ:

1) $E_x = E_t E_e (x \mid t)$,
where $E_e (x \mid t)$ is the expectation with fixed but arbitrary t taken with respect to the distribution of e and E_t the expectation of the result with respect to the distribution of t;
2) $\Sigma(x) = E_t (\Sigma_e(x \mid t)) + \Sigma_t (E_e(x \mid t))$, where $\Sigma(x \mid t)$ is the covariance matrix of x with fixed but arbitrary t, taken with respect to the distribution of e, resulting in a function of t, whereupon the expectation E_t of this function is taken with respect to the distribution of t, while the second term is the covariance matrix with respect to the distribution of t, calculated over the conditional expectation $E_e (x \mid t)$.
It follows from 1) that: $Ex = E_t E_e (x \mid t) = E_t E_e (t + e \mid t) = E_t t = \mu_t$, a vector of means.
2) consists of two terms: $\Sigma_e(t + e \mid t) = \Sigma_e(e \mid t)$.
This is a diagonal matrix, because e_i and e_j are mutually independent for any value of t. On the diagonal $\sigma^2 (e_i \mid t)$ stands for i = 1, 2, ..., k.

* T denotes transposition of a vector.

$E_e(t + e|t) = t$ and $\Sigma_t(t)$ is a matrix with cov (t_i, t_j) outside the diagonal and $\sigma^2(t_i)$ on the diagonal.

The correlation coefficient $\rho\ (x_i, x_j)$ can be written thus:

$$\rho(x_i, x_j) = \frac{\mathrm{cov}(t_i, t_j)}{\sigma(x_i)\ \sigma(x_j)}$$

Further, $t_i = \lambda w_i$. Thus:

$$\sigma^2(t_i) = w^2_i \sigma^2\ (\lambda);\ \mathrm{cov}\ (t_i, t_j) = w_i w_j\ \sigma^2(\lambda) = \sigma(t_i)\sigma(t_j).$$

Hence,

$$\rho\ (x_i, x_j) = \frac{\sigma(t_i)}{\sigma(x_i)} \cdot \frac{\sigma(t_j)}{\sigma(x_j)}$$

In addition, we need - for the discussion in section 8 - the correlation $\rho(t_i, x_i)$. Now cov $(t_i, x_i) = \sigma^2(t_i) + \mathrm{cov}\ (t_i, e_i) = \sigma^2(t_i)$ and thus:

$$\rho(t_i, x_i) = \frac{\sigma(t_i)}{\sigma(x_i)}.$$

Comparison with the expression for $\rho(x_i, x_j)$ reveals that

$$\rho(x_i, x_j) = \rho(t_i, x_i) \times \rho(t_j, x_j).$$

If $\rho(t_i, x_i)$ is reasonably constant for different i, $\rho(t_i, x_i)$ is approximately equal to $\sqrt{\rho(x_i, x_j)}$.

Hitherto, we have not made use of the Poisson properties, but only of the notation $t_i = \lambda w_i$.

We shall now reconsider the diagonal of $\Sigma(x)$, i.e. the variances $\sigma^2(x_i)$.

These can be written as follows:

$$\sigma^2(x_i) = E_t \sigma^2(e_i|t) + \sigma^2(t_i).$$

For the Poisson model:

$$\sigma^2(e_i|t) = t_i, \text{ thus } \sigma^2(x_i) = Et_i + \sigma^2(t_i).$$

Hence, if the Poisson model applies,

$$\rho(x_i, x_j) = \left\{ \frac{\sigma^2(x_i) - Et_i}{\sigma^2(x_i)} \cdot \frac{\sigma^2(x_j) - Et_j}{\sigma^2(x_j)} \right\}^{\frac{1}{2}}$$

$Et_i = \mu_{t_i} = E(x_i)$ and thus $\rho(x_i, x_j)$ can be estimated by substituting $s^2(x_i)$ for $\sigma^2(x_i)$ and $\bar{x}_i$ for μ_{t_i}. For $\rho(t_i, x_i)$ we find

$$\left\{ \frac{\sigma^2(x_i) - Et_i}{\sigma^2(x_i)} \right\}^{\frac{1}{2}}$$

and these values can also be estimated, with the same substitutions as for $\rho(x_i, x_j)$.

REFERENCES

CHADWICK-JONES, G.K., C.A. BROWN, N. NICKOLSON & C. SHEPPARD. Absence measures: their reliability and stability in an industrial setting. Personnel Psychology 24 (1971) 463-70

DE COCK, G., & F. CORTHOUTS. Psychologische aspekten van het vraagstuk van de ongevallenpreventie. In: Burger, G.C.E. (red.), Arbeids- en bedrijfsgeneeskunde, Leiden, 1974

FARR, J.L., B.S. O'LEARY & C.J. BARTLETT. Ethnic group membership as a moderator of the prediction of job performance. Personnel Psychology 24 (1971) 609-36

FELLER, W. An introduction to probability theory and its applications, vol. 1, New York, John Wiley, 1950

FERGUSON, D. Some characteristics of repeated sickness absence. Brit. J.industr.Med. 29 (1972) 420-31

FROGGATT, P. Short-term absence from industry III; the inference of proneness and a search for causes. Brit.J.industr.Med. 27 (1970) 297-312

HINKLE, L.E., R.H. PINSKY, I.D.J. BROSS & N. PLUMMER. The distribution of sickness disability in a homogeneous group of 'healthy adult men'. Amer.J.Hyg. 64 (1956) 220-42

HINKLE, L.E., N. PLUMMER & L. HOLLAND WHITNEY. The continuity of patterns of illness and the prediction of future health. J.occup. Med. 3 (1961) 417-23

HUSE, E.F., & E.K. TAYLOR. The reliability of absence measures. J.appl. Psychology 46 (1962) 159-60

INDEX OF HEALTH; mathematical models. Nat.Center Hlth.Stat. (1965) (series 2 nr. 5)

LATHAM, G.P., & E.D. PURSELL. Measuring absenteeism from the opposite side of the coin. J.appl.Psychology 60 (1975) 369-71

LYONS, T.F. Turnover and absenteeism: a review of relationships and shared correlates. Personnel Psychology 55 (1972) 271-81

MORGAN, L.G., & J.B. HERMAN. Perceived consequences of absenteeism. J.appl.Psychology 61 (1976) 738-42

MUCHINSKY, P.M. Employee absenteeism: a review of the literature. J.vocat.behavior 10 (1977) 316-40

NASS, Ch. Verdelingsvrije toetsen voor verschillen in verzuimfrekwentie. Voordracht gehouden op de biometrische dag, Utrecht, 18 mei 1956

NICHOLSON, N., & P.M. GOODGE. The influence of social organisational and biographical factors on female absence. J.management studies 13 (1976) 234-54

PHILIPSEN, H. Levensfase, generatie en ziekteverzuim. Maastricht, Rijksuniversiteit Limburg/NIPG/TNO, 1977

PLACKETT, R.L. The analysis of categorical data. London, Griffin, 1974

SHAW, L., & H. SICHEL. Accident proneness research in the occurrence, causation and prevention of road accidents. Oxford, Pergamon Press, 1971

TURNER, W.W. Dimensions of foreman performance: a factor analysis of criterion measures. J.appl.Psychology 44 (1960) 216-23

WATERS, L.K., & D. ROACH. Relationship between job attitudes and two forms of withdrawal from the work situation. J.appl.Psychology 55 (1971) 92-4

WATERS, L.K., & D. ROACH. Job satisfaction, behavioral intention, and absenteeism as predictors of turnover. Personnel Psychology 32 (1979) 393-7