Sickness absence rates are higher in winter than in summer and also the incidence of many diseases is affected by climatic conditions. Both these facts lead one to suspect that there is a genuine causative relationship between the weather and sickness absence. The object of this investigation is to develop a model to illustrate this relationship in order to obtain some idea of its precise form.

Any effect of weather on the weekly incidence of sickness absence can be assumed to take one of three forms:

(i) a steady long-term cumulative effect whereby the more adverse weather conditions in winter eventually affect the health and absence behaviour of employees;

(ii) a more short-term effect whereby deviations of weather conditions about the seasonal norm may influence the sickness absence frequency in the same or subsequent weeks;

(iii) the spread of infectious disease, whereby one week's sickness absence frequency is determined to some extent by the amount of infection brought over from previous weeks. Though waves of infection are influenced by weather conditions (e.g., they are more frequent in winter), their occurrence is largely unpredictable and they must be considered separately from (i) and (ii).

The regression model described below is designed to assess the importance of the effects of types (ii) and (iii) after the elimination of the broad seasonal trends indicative of effects of type (i). This model is illustrated by sickness absence data from one organization.

**LITERATURE**

Little research has been carried out on the effect of the weather on absenteeism. However, several attempts have been made to discover the relationship between meteorological factors and various indices of mortality and morbidity, in most cases for particular types of disease.

**MORTALITY STUDIES**

Young (1924), Russell (1926), and Wright and Wright (1945) are three early studies of the effect of climate on weekly or monthly numbers of deaths due to respiratory disease. Their analyses based on simple correlational techniques showed that lower temperatures were associated with higher mortality rates. More recently, Boyd (1960) looked at the association between various measures of climate and air pollution and deaths from respiratory and heart disease in London and East Anglia during eight winters. Temperature and humidity two weeks before death had strongest correlations with weekly mortality whereas pollution was of secondary importance. The effect of fog appeared only during periods of low temperature and then for respiratory disease only.

One serious flaw in all these studies is the failure to take into account the infectious nature of respiratory disease. For instance, during an influenza epidemic the weekly deaths from respiratory disease relate more directly to the amount of infection in previous weeks than to the actual weather conditions, one measure of this infection being the previous week's numbers of deaths. Thus, the autocorrelation present in such time series of respiratory deaths must be taken into account before one can comment on the effect of weather conditions. One further fault in the methods of Wright and Wright and Boyd is the failure to eliminate broad seasonal trends (e.g., by using deviations from a moving average or some parametric trend) before investigating correlations between disease and meteorological factors. Spicer (1959) draws attention to the need to use such deviations, sometimes called residuals, and this approach has been adopted in the regression model described below.

**MORBIDITY STUDIES**

Holland, Spicer, and Wilson (1961) used two measures of the monthly occurrence of disease—admissions to the Emergency Bed Service of a London hospital and admissions to the sick quarters in three Royal Air Force stations. The climatic factors considered were average temperature, absolute humidity, rainfall, barometric pressure, sun, and pollution (smoke), and analysis of covariance was used to eliminate broad seasonal and
annual trends. For the London hospital there was no significant relationship between non-respiratory disease and the weather. However, for respiratory illness, temperature and air pollution both had a significant effect in adults only. For the RAF stations, the effect of temperature on respiratory illness was evident.

Davey and Reid (1972) studied the relationship between air temperature and outbreaks of influenza using a straightforward descriptive approach which perhaps makes it difficult to generalize from their conclusions.

The effect of climate on the incidence of poliomyelitis has provoked a considerable amount of research, much of which is critically summarized in Lawrence (1962). The incidence of poliomyelitis was higher in summer but the factor or combination of factors (climatic or other) which caused this summer peak was not immediately obvious. There are multitudinous variables with seasonal cycles and their close association makes distinction between them difficult. Furthermore, Lawrence points out that the variations in meteorological conditions over short periods and over quite small areas cannot be adequately illustrated in standard meteorological data.

Spicer (1959) and Wise (1966) illustrate the type of methods most appropriate for estimating the effect of climate on disease. They both use the same data consisting of the monthly numbers of poliomyelitis inceptions during 1947-56 with the corresponding averages for temperature and relative humidity. Spicer used an analysis of covariance model to remove annual and seasonal trends, thus obtaining multiple regression coefficients of log (poliomyelitis incidence) on the climatic variables in the same and the preceding two months. All six coefficients were bigger than their standard errors (positive for temperature and negative for relative humidity) with temperature in the preceding month making the greatest contribution, and Spicer thus inferred that warm, dry weather does play a part in the spread of the disease.

Wise first points out that the main factor determining one month's poliomyelitis rate is the previous month's incidence of the disease. He thus elaborates on Spicer's model by introducing the previous month's poliomyelitis as an extra covariate. He also allows the regression coefficients for temperature and relative humidity to have different values for different calendar months; the consequent peak effects are shown to occur in late summer and early spring. He concludes that the temperatures in July-September have a direct influence on poliomyelitis occurrence through to October and perhaps even December. Wise's analysis thus provided greater insight into the behavioural pattern of the disease.

DATA AND METHODS

SICKNESS ABSENCE

One limitation in many investigations of the relationship between weather and disease is that the data about the disease characteristic are often collected over large areas (e.g., national or regional figures) with the result that local variations in meteorological conditions or disease incidence cannot be taken into account. Such a macroscopic approach can result in failure to discover the true nature of such a relationship. A similar failure can occur if data are grouped according to quite lengthy time intervals (e.g., a month or more). Fortunately, sickness absence is a sufficiently common phenomenon for one to be able to adopt a more detailed approach, the records of a single factory being ideal for this purpose. In this case, one site of a large continuous process industry in south-east England has been chosen.

The population under study is defined as the 1,057 male manual workers who were in employment from 1960 to 1964. The numbers of spells of sickness absence starting in each week of this five-year period have been accumulated from the records of individual employees. Sickness absence includes all spells of absence due to medical incapacity (and thus accompanied by a doctor's certificate) which lasted for one shift or more. Self-certification of short spells was introduced in 1965 and the subsequent change in absence behaviour means that the period of observation is before then. Diagnostic information has enabled the weekly numbers of spells to be categorized into upper respiratory disease, bronchitis, and non-respiratory illness.

The estimated seasonal trend in the total weekly spells of sickness absence is shown as a continuous line in Figure 1. This curve has been obtained by using harmonic analysis and the details of this approach are given by Pocock (1972). The dotted line in Figure 1 is a simple harmonic of period one year and can be seen to provide a good fit to the seasonal trend except for the peak at New Year, the drop in absence before Christmas, and the slight drop during August.

It is essential to allow for the effects of non-climatic factors on sickness absence before studying its association with the weather, and these exceptions indicate that one such factor is the occurrence of annual and public holidays. At the factory annual holidays were spread throughout the summer but the peak holiday period was from mid-July to mid-August and the relevant six weeks of each year have
been excluded from analysis. The effect of Easter and Whitsun on sickness absence is short-lived so that the exclusion of one week (Thursday to Wednesday) was assumed to be adequate in each case. The effect of Christmas and New Year is more long-term and it is necessary to exclude a three-week period starting the week before Christmas.

The occurrence of an influenza epidemic causes such an extreme rise in sickness absence frequency that it might be appropriate to exclude the relevant weeks. However, in 1960-64 the only severe epidemic occurred during the Christmas period and the problem is thus avoided.

There was no evidence of any long-term trends in sickness absence over this period, the annual numbers of spells being fairly stable.

The weekly intervals are taken from Thursday to Wednesday since this enables the correction for Easter to be restricted to one week only. As shift workers are involved in weekend working and form over one third of the total work force, there is no obvious dividing point for the week, so that it is valid to take the most convenient division.

The period 1960-64 contains a total of 261 weeks (including one day from 1959 and excluding one from 1964 so that the week starts on Thursday). However, the exclusion of holidays reduced the number of weeks in the regression analysis to 186. The average spells per week of sickness absence during this period was 22.9 with standard deviation 7.08.

To dampen the greater variations encountered in winter absence frequencies a square-root transformation has been applied to all weekly numbers of spells. The theoretical basis for this is that if weekly absence frequency is a Poisson random variable with varying mean, its square-root has approximately constant variance equal to \(\frac{1}{\lambda}\).

### The Weather

The nearest weather station to the factory was 10 miles away, which should be near enough to make any differences in weather between the two places negligible. The following meteorological information was available for every day from 1959 to 1964:

1. the maximum temperature (MAX TEMP) for the 24-hour period starting at 9 a.m. on the day, to an accuracy of 1°F;
2. the minimum temperature (MIN TEMP) for the 24-hour period starting at 9 a.m. the previous day, to an accuracy of 1°F;
3. the rainfall (RAIN) recorded in \(\frac{1}{10}\)ths of an inch for the 24 hours starting at 9 a.m. on the day;
4. the hours of sunshine (SUN) recorded to an accuracy of \(\frac{1}{10}\)th of an hour for the whole day.

It would have been desirable to obtain details of barometric pressure and relative humidity but these were not available. However, the high correlations between meteorological factors, as shown by Boyd (1960), make it difficult to distinguish between the effects of some variables, and thus the inclusion of more might merely confuse the conclusions from the analysis. The weekly averages of the four measurements, MAX TEMP, MIN TEMP, RAIN, and SUN, are the meteorological variables used in the model.

The weather back as far as eight weeks before absence is used so that a few week's data from 1959 were needed.

To illustrate seasonal variations in meteorological factors the monthly averages of mean temperature, total rainfall, and total hours of sun for the above-mentioned weather station for 1921-50 are shown in Figure 2; although these years do not include the period of the survey, the use of 30-year averages gives a more reliable estimate of the seasonal trends. For mean temperature and hours of sun the smooth trends obtained should be adequately represented by harmonic curves. For rainfall the seasonal curve is more irregular but the large standard deviations about this mean curve, not shown here but present in all British rainfall data, make the effect of the trend itself of lesser importance. Thus, it is assumed that the elimination of trend in the four meteorological variables mentioned above can be adequately obtained by use of the appropriate harmonic curves of period one year.

### The Regression Model

The objective here is to describe a method for analysing variations in weekly sickness absence...

---

**Figure 1.** Estimated seasonal trend in weekly spells (1960-64).
frequency over a number of years $N$ with a total number of weeks $K$.

Suppose there are $J$ types of absence (e.g., diagnostic groups) and define $Y_{ij}$ as the square-root of the number of spells of type $j$ starting in week $i$, for $j = 1, \ldots, J$ and $i = 1, \ldots, K$. These $J$ types are not necessarily exclusive or exhaustive.

Suppose there are $H$ meteorological factors with their weekly averages $W_{ih}$ for $h = 1, \ldots, H$ and $i = 1, \ldots, K$.

Now consider absence of a particular type $k$ and the combination of factors which determine its weekly values $Y_{ik}$. There are five such factors:

(i) a broad seasonal trend, characterized by a harmonic of period one year such that $Y_{ik}$ has expected value $a_o + b_o \cos \left( \frac{2\pi Ni}{K} + \alpha_o \right)$ before modification for the effects of (ii) - (v) below, where $a_o$, $b_o$, and $\alpha_o$ are constants.

(ii) deviations of weather conditions about the seasonal norm for week $i$ and the $S$ preceding weeks. Since the seasonal trends for these $H$ factors are assumed to be harmonic, such deviations can be defined as

$$\{ W_{i-s,h} - b_{sh} \cos \left( \frac{2\pi Ni}{K} + \beta_{sh} \right) \}$$

for $s = 0, \ldots, S$ and $h = 1, \ldots, H$ where $\{b_{sh}\}$ and $\{\beta_{sh}\}$ are constants;

(iii) deviations of previous weeks’ sickness absence frequencies about the seasonal norm. Considering up to $T$ preceding weeks and assuming seasonal trends to be harmonic, such deviations can be defined as

$$\{ Y_{i-t,j} - a_{ij} \cos \left( \frac{2\pi Ni}{K} + \alpha_{ij} \right) \}$$

for $t = 1, \ldots, T$ and $j = 1, \ldots, J$ where $\{a_{ij}\}$ and $\{\alpha_{ij}\}$ are constants;

The effect of such deviations on $Y_{ik}$ can be attributed to the carry-over of infections from one week to the next;

(iv) other unknown factors which are assumed to be sufficiently unsystematic to be treated as random variations;

(v) the occurrence of holidays. This factor has been eliminated by excluding the relevant weeks from the data and it is therefore not included henceforth.

Under the assumption that the effects of (i) to (iv) are additive, the following linear regression model is obtained:

$$Y_{ik} = c + a \cos \frac{2\pi Ni}{K} + b \sin \frac{2\pi Ni}{K}$$

$$+ \sum_{h=1}^{H} \sum_{s=o}^{S} d_{sh} W_{i-s,h}$$

$$+ \sum_{j=1}^{J} \sum_{t=1}^{T} e_{ij} Y_{i-t,j} + e_i$$

where $c$, $a$, $b$, $\{d_{sh}\}$ and $\{e_{ij}\}$ are regression coefficients and $\{e_i\}$ are independent, identically distributed random variables with zero mean. The harmonic terms in this equation represent the addition of the harmonic terms in (i) to (iii).

Estimation Procedure

Least squares is the method to be used for estimation of the regression coefficients, but the fitting of the full regression model is liable to give rise to a large number of non-significant estimates. Thus, the method of stepwise regression, as described
by Efroymson (1960), is adopted, introducing variables into the regression model one by one until there exist no further variables with coefficients significantly different from zero. One modification to Efroymson’s method is that the two harmonic terms are included as ‘forced variables’, i.e., the coefficients $a$ and $b$ are included as the first two steps in the regression and cannot be excluded. This is essential as the model is based on the investigation of variations about the seasonal trend.

**Results**

The regression model is now applied to the data described earlier. The period covered is 1960-64 so that $N = 5$ years and $K = 261$ weeks.

The following four types of absence are considered (so that $J = 4$): $j = 1$ represents all spells of sickness absence, $j = 2$ spells due to upper respiratory disease, $j = 3$ spells of bronchitis, and $j = 4$ spells diagnosed as non-respiratory.

It seems reasonable to suppose that the detection of any possible relationship between absence frequencies in adjacent weeks may be assisted by the use of diagnostic information. Furthermore, the lack of seasonal trends in all non-respiratory absence makes it sensible to combine all such absence into one type. $k$ is set equal to 1 so that the model is applied with total weekly spells as the dependent variable.

Also, $T = 2$, so that the effect of absence up to two weeks previously is considered. The exclusion of all weeks which contain holiday period absence as either dependent or independent variables means that the number of weeks for analysis is reduced to 166.

Meteorological data are as described earlier so that $H = 4$ and $h = 1, 2, 3$, and 4 respectively for MAX TEMP, MIN TEMP, RAIN, and SUN. $S = 8$, so that the effect of weather up to eight weeks previously is considered.

A first analysis showed that the absence frequencies two weeks previously did not have significant coefficients. Consequently, analysis was repeated with $T = 1$, thus enabling the number of weeks to be increased to 186.

The results of this second analysis are summarized in Table I. Variables are listed in the order in which they entered the model with their estimated regression coefficients and standard errors in the final model.

Thus, it can be seen that the previous week’s upper respiratory absence and the mean maximum temperature two weeks previously are both significant (at the 1% level) in determining one week’s total frequency of absence, after elimination of seasonal trends. Furthermore, no other variable was significant at the 5% level. In particular, it is interesting that respiratory absence in the previous week takes precedence over total absence for inclusion in the model.

The following analysis of variance table is obtained from the model with the four independent variables presented in Table I.

Thus, 51% of the variation of $Y_{11}$ can be explained by the regression.

The residual mean square is 0.268, close to the value 1/4 to be expected if all non-random variation is contained in the regression. This is evidence that the model adequately explains weekly variations in sickness absence.

It is unsatisfactory at this stage to assume that the mean maximum temperature only affects absence two weeks before the event. As shown by Wise (1966), it is sensible to consider a weighted sum of the mean maximum temperatures over a period of several weeks before absence, this being done by introducing the relevant variables into the regression model.

Thus, MAX TEMP 0, 1, and 3 weeks before absence are included in a new expanded regression model resulting in the regression coefficients listed in Table II.

The magnitudes of the regression coefficients for mean maximum temperature 0, 1, and 3 weeks before absence are all less than half their respective
standard errors, indicating that the inverse relationship between maximum temperature and total absence exists only when considering the temperature 2 weeks before absence.

The equivalent regression model using mean minimum temperature instead of mean maximum temperature resulted in a slightly smaller regression sum of squares, indicating a preference for maximum temperature in the model.

So far the effects of weather and previous absence on weekly total absence have been assumed to be the same for all times of year. To check on the validity of this assumption the 186 weeks are divided into three times of year as follows:

<table>
<thead>
<tr>
<th>Time of Year</th>
<th>No. of Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>January-March (winter)</td>
<td>55</td>
</tr>
<tr>
<td>April-June (spring)</td>
<td>55</td>
</tr>
<tr>
<td>Late August-December (autumn)</td>
<td>76</td>
</tr>
</tbody>
</table>

It would have been ideal to break the year into four seasons but the annual holidays make analysis of summer absenteeism impracticable. To ensure whole numbers of weeks for ‘winter’ and ‘spring’ the dividing point varied slightly between years.

Separate regressions have been performed for the three times of year, attention being restricted to those independent variables found significant in the overall regression model above. That is, Cos $\frac{2\pi Ni}{K}$, Sin $\frac{2\pi Ni}{K}$, $Y_{i-1,2}$, and $W_{i-2,1}$ are the four independent variables. The results of these three regressions are shown in Table III.

$Y_{i-1,2}$ is the square root of the number of spells of upper respiratory disease for the week before the one under consideration. From its three regression coefficients for the three ‘seasons’ it can be seen that the effect of previous upper respiratory absence on weekly total absence is most marked in the winter months. However, the inverse relationship between weekly total absence frequency and temperature two weeks previously is more consistent for all times of year, though significance is achieved only in winter.

In order to check the validity of a regression model an analysis of the residuals (i.e., differences between observed and expected weekly spells) is customary. No systematic deviations were found except for the winters of 1963 and 1964. For the former, 9 out of 11 residuals were positive, whereas the latter had all 11 residuals negative. The winter of 1963 was extremely severe and it could be that the effect of long periods of very cold weather had a more serious effect on sickness absence than the model estimated. The lower than expected absence rates in early 1964 might be due to the fear of redundancies present at that time, though this could not be detected in the absence rate for the whole year.

Pocock (1972) showed that non-random variation in weekly sickness absence frequency is marked only for upper respiratory disease and bronchitis. Therefore it is reasonable to assume that the relationships discovered in this investigation are predominantly due to respiratory illness. Furthermore, since upper respiratory disease occurs more frequently than bronchitis, this last comment applies mainly to the former absence category.

However, both non-respiratory illness and bronchitis can be studied separately. First, using the notation defined above, the correlation coefficient for $Y_{i4}$ and $Y_{i-1,2}$ over the 186 weeks used in the above model has a value 0.182 ($P < 0.05$). This is evidence that one week’s frequency of non-respiratory absence may be affected by the incidence of upper respiratory disease in the previous week.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Winter</th>
<th>Spring</th>
<th>Autumn</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{2\pi Ni}{K}$</td>
<td>-0.111</td>
<td>0.015</td>
<td>0.821</td>
</tr>
<tr>
<td>Sin $\frac{2\pi Ni}{K}$</td>
<td>-0.027</td>
<td>0.043</td>
<td>-0.683</td>
</tr>
<tr>
<td>$Y_{i-1,2}$</td>
<td>0.396†</td>
<td>0.045</td>
<td>0.152</td>
</tr>
<tr>
<td>$W_{i-2,1}$</td>
<td>-0.023*</td>
<td>0.024</td>
<td>-0.015</td>
</tr>
</tbody>
</table>

*$P < 0.05$
†$P < 0.01$
indicating that bouts of respiratory infection may lead to a worsening in the general health of a population.

Suppose $Y_{13}$, i.e., weekly frequency of bronchitis, is the dependent variable in the regression equation and the model is applied to the 55 winter weeks of 1960-64. Then, after the removal of seasonal trends, $W_{1-1,1}$ is the only significant variable in the stepwise regression, it being significant at the 1% level. Thus, the incidence of absence due to bronchitis is most closely related to the temperature in the previous week. Boyd (1960) showed that mortality due to bronchitis was most affected by temperature two weeks before death. However, these two results are not conflicting since absence is a more minor manifestation of the disease process which one might expect to be more immediately affected by meteorological factors.

**DISCUSSION**

Tromp (1964) gives a general description of the methods of bio-meteorological analysis. He cites two main approaches: the *empirical method* where one is attempting to define the form of the weather-disease relationship and the *experimental approach* where one is concerned with validating more specific theories or hypotheses. The former necessarily precedes the latter since one must have some idea of the problem before one can produce elaborate theories.

The approach adopted in this investigation has been 'empirical'. This enables one to suggest the relationship between the weather and various aspects of sickness absence, but one is not able to explain these relationships without further experimental research. However, the results of this study should be a guide as to the direction which future research could take.

Tromp gives two general ways in which the weather might affect the occurrence of disease:

(i) a *physiological* effect, whereby the resistance of the body to disease is lowered;

(ii) a *social* effect, whereby the changed social behaviour (e.g., crowding in rooms) increases the spread of infection.

When looking at sickness absence one must remember that it is not merely a measure of disease but also of the motivation to go to work, and this introduces a third possible way in which the weather might play a part:

(iii) a *motivational* effect, whereby weather conditions affect the employee’s ‘will to work’.

This study has shown that the weekly frequency of absence is related to two factors after the elimination of seasonal trends:

(i) the previous week’s frequency of absence due to upper respiratory disease, this effect being strongest in winter;

(ii) the mean maximum temperature two weeks before, which seemed to have a slight effect in all seasons, was again most marked in winter.

Overall, (i) was more important than (ii), which implies that the spread of respiratory infection plays a dominant role in determining weekly sickness absence frequency. There is no evidence that weather conditions in the current week affected the weekly absence frequency, and therefore the postulated motivational effect mentioned above would appear to be of minor importance, variations in weekly absence frequency being largely due to variations in the incidence of disease. This negative finding is of some interest in view of the widely accepted view that personal motivation, and job satisfaction in particular, plays an important part in the genesis of a short spell of sickness absence (Office of Health Economics, 1971).

This investigation is essentially dealing with weekly fluctuations in the *total* incidence of sickness absence, and the model is therefore not designed for the detailed study of epidemics of particular diseases. Nevertheless, an outbreak of influenza on a large scale might be thought to build up sufficient immunity in the community actually to cause the incidence of respiratory illness (and therefore all illness) to fall below average in the aftermath of such an epidemic. Evidence in support of this can be found in the national weekly claims for sickness benefit for 1969-70 (published by the Department of Health and Social Security, 1971). The influenza outbreak in December/January was followed by below average sickness not only in the week immediately after the epidemic but for the whole of February and March. It is possible, therefore, that the model might be somewhat inadequate in such an extreme situation. However, in most years (and certainly during the period used in the results) an epidemic either does not occur or is much less severe. For example, in the population of 1,057 men, never more than 40 went off sick with respiratory disease in any one week and, therefore, there would appear to be little chance of widespread immunity arising.

One interesting future project would be to investigate whether the spread of respiratory infection illustrated in this study occurred primarily in the work or in the home environment. This could perhaps be achieved by the study of space-time clustering in the incidence of respiratory absence with two space factors, the employee’s location at work and his location at home.
As regards the effect of temperature on sickness absence, it might be thought that the difference in temperature between one week and the next is more important than the actual level. For instance, a sharp fall in temperature might lead to increased sickness absence, whereas people might become accustomed to a continuing cold spell. However, if such a relationship did occur it would have been detected in the estimation of the regression equation shown in Table II.

**Summary**

A regression model is developed for investigating the effect of weather and respiratory infection on the incidence of sickness absence in an industrial population. After elimination of broad seasonal trends, the number of spells of upper respiratory disease in the previous week and the temperature two weeks previously are found to be the most important factors determining sickness absence in any one week.

I am very grateful to Professor P. Armitage and Dr. P. J. Taylor for their invaluable advice throughout this investigation. I would also like to thank the Company for permission to use their sickness absence records.

**References**


*Author’s present address: Statistical Laboratory, State University of New York at Buffalo, Amherst, New York 14226, U.S.A.*